Springer Finance

Springer
New York
Berlin
Heidelberg
Hong Kong
London
Milan
Paris
Tokyo

Springer Finance

Springer Finance is a programme of books aimed at students, academics, and practitioners working on increasingly technical approaches to the analysis of financial markets. It aims to cover a variety of topics, not only mathematical finance but foreign exchanges, term structure, risk management, portfolio theory, equity derivatives, and financial economics.

M. Ammann, Credit Risk Valuation: Methods, Models, and Applications (2001)
E. Barucci, Financial Markets Theory: Equilibrium, Efficiency and Information (2003)
N.H. Bingham and R. Kiesel, Risk-Neutral Valuation: Pricing and Hedging of Financial Derivatives, 2nd Edition (2004)
T.R. Bielecki and M. Rutkowski, Credit Risk: Modeling, Valuation and Hedging (2001)
D. Brigo amd F. Mercurio, Interest Rate Models: Theory and Practice (2001)
R. Buff, Uncertain Volatility Models – Theory and Application (2002)
R.-A. Dana and M. Jeanblanc, Financial Markets in Continuous Time (2003)
G. Deboeck and T. Kohonen (Editors), Visual Explorations in Finance with Self-Organizing Maps (1998)
R.J. Elliott and P.E. Kopp, Mathematics of Financial Markets (1999)
H. Geman, D. Madan, S.R. Pliska and T. Vorst (Editors), Mathematical Finance – Bachelier Congress 2000 (2001)
M. Gundlach and F. Lehrbass (Editors), CreditRisk+ in the Banking Industry (2004)
Y.-K. Kwok, Mathematical Models of Financial Derivatives (1998)
M. Külpmann, Irrational Exuberance Reconsidered: The Cross Section of Stock Returns, 2nd Edition (2004)
A. Pelsser, Efficient Methods for Valuing Interest Rate Derivatives (2000)
J.-L. Prigent, Weak Convergence of Financial Markets (2003)
B. Schmid, Credit Risk Pricing Models: Theory and Practice, 2nd Edition (2004)
S.E. Shreve, Stochastic Calculus for Finance I: The Binomial Asset Pricing Model (2004)
S.E. Shreve, Stochastic Calculus for Finance II: Continuous-Time Models (2004)
M. Yor, Exponential Functionals of Brownian Motion and Related Processes (2001)
R. Zagst, Interest-Rate Management (2002)
Y.-l. Zhu and I.-L Chern, Derivative Securities and Difference Methods (2004)
A. Ziegler, Incomplete Information and Heterogeneous Beliefs in Continuous-Time Finance (2003)
A. Ziegler, A Game Theory Analysis of Options: Corporate Finance and Financial Intermediation in Continuous Time, 2nd Edition (2004)

Steven E. Shreve

Stochastic Calculus for Finance I

The Binomial Asset Pricing Model

With 33 Figures

Springer

Steven E. Shreve
Department of Mathematical Sciences
Carnegie Mellon University
Pittsburgh, PA 15213
USA
shreve@cmu.edu

Mathematics Subject Classification (2000): 60-01, 60H10, 60J65, 91B28

Library of Congress Cataloging-in-Publication Data
Shreve, Steven E.
Stochastic calculus for finance / Steven E. Shreve.
p. cm. — (Springer finance series)
Includes bibliographical references and index.
Contents v. 1. The binomial asset pricing model.
ISBN 0-387-40100-8 (alk. paper)
1. Finance—Mathematical models—Textbooks. 2. Stochastic analysis—Textbooks. I. Title. II. Springer finance.
HG106.S57 2003
332'.01'51922—dc22 2003063342

ISBN 0-387-40100-8 Printed on acid-free paper.

Printed in the United States of America. (AL/MVY)

9 8 7 6 5 4 3 2 1 SPIN 10929445

Springer-Verlag is a part of *Springer Science+Business Media*

springeronline.com

To my students

Preface

Origin of This Text

This text has evolved from mathematics courses in the Master of Science in Computational Finance (MSCF) program at Carnegie Mellon University. The content of this book has been used successfully with students whose mathematics background consists of calculus and calculus-based probability. The text gives precise statements of results, plausibility arguments, and even some proofs, but more importantly, intuitive explanations developed and refined through classroom experience with this material are provided. Exercises conclude every chapter. Some of these extend the theory and others are drawn from practical problems in quantitative finance.

The first three chapters of Volume I have been used in a half-semester course in the MSCF program. The full Volume I has been used in a full-semester course in the Carnegie Mellon Bachelor's program in Computational Finance. Volume II was developed to support three half-semester courses in the MSCF program.

Dedication

Since its inception in 1994, the Carnegie Mellon Master's program in Computational Finance has graduated hundreds of students. These people, who have come from a variety of educational and professional backgrounds, have been a joy to teach. They have been eager to learn, asking questions that stimulated thinking, working hard to understand the material both theoretically and practically, and often requesting the inclusion of additional topics. Many came from the finance industry, and were gracious in sharing their knowledge in ways that enhanced the classroom experience for all.

This text and my own store of knowledge have benefited greatly from interactions with the MSCF students, and I continue to learn from the MSCF alumni. I take this opportunity to express gratitude to these students and former students by dedicating this work to them.

Acknowledgments

Conversations with several people, including my colleagues David Heath and Dmitry Kramkov, have influenced this text. Łukasz Kruk read much of the manuscript and provided numerous comments and corrections. Other students and faculty have pointed out errors in and suggested improvements of earlier drafts of this work. Some of these are Jonathan Anderson, Bogdan Doytchinov, Steven Gillispie, Sean Jones, Anatoli Karolik, Andrzej Krause, Petr Luksan, Sergey Myagchilov, Nicki Rasmussen, Isaac Sonin, Massimo Tassan-Solet, David Whitaker and Uwe Wystup. In some cases, users of these earlier drafts have suggested exercises or examples, and their contributions are acknowledged at appropriate points in the text. To all those who aided in the development of this text, I am most grateful.

During the creation of this text, the author was partially supported by the National Science Foundation under grants DMS-9802464, DMS-0103814, and DMS-0139911. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the author and do not necessarily reflect the views of the National Science Foundation.

Pittsburgh, Pennsylvania, USA
December 2003

Steven E. Shreve

Contents

Introduction

Background

By awarding Harry Markowitz, William Sharpe, and Merton Miller the 1990 Nobel Prize in Economics, the Nobel Prize Committee brought to worldwide attention the fact that the previous forty years had seen the emergence of a new scientific discipline, the "theory of finance." This theory attempts to understand how financial markets work, how to make them more efficient, and how they should be regulated. It explains and enhances the important role these markets play in capital allocation and risk reduction to facilitate economic activity. Without losing its application to practical aspects of trading and regulation, the theory of finance has become increasingly mathematical, to the point that problems in finance are now driving research in mathematics.

Harry Markowitz's 1952 Ph.D. thesis *Portfolio Selection* laid the groundwork for the mathematical theory of finance. Markowitz developed a notion of mean return and covariances for common stocks that allowed him to quantify the concept of "diversification" in a market. He showed how to compute the mean return and variance for a given portfolio and argued that investors should hold only those portfolios whose variance is minimal among all portfolios with a given mean return. Although the language of finance now involves stochastic (Itô) calculus, management of risk in a quantifiable manner is the underlying theme of the modern theory and practice of quantitative finance.

In 1969, Robert Merton introduced stochastic calculus into the study of finance. Merton was motivated by the desire to understand how prices are set in financial markets, which is the classical economics question of "equilibrium," and in later papers he used the machinery of stochastic calculus to begin investigation of this issue.

At the same time as Merton's work and with Merton's assistance, Fischer Black and Myron Scholes were developing their celebrated option pricing formula. This work won the 1997 Nobel Prize in Economics. It provided a satisfying solution to an important practical problem, that of finding a fair price for a European call option (i.e., the right to buy one share of a given stock at a specified price and time). In the period 1979–1983, Harrison, Kreps, and Pliska used the general theory of continuous-time stochastic processes to put the Black-Scholes option-pricing formula on a solid theoretical basis, and, as a result, showed how to price numerous other "derivative" securities.

Many of the theoretical developments in finance have found immediate application in financial markets. To understand how they are applied, we digress for a moment on the role of financial institutions. A principal function of a nation's financial institutions is to act as a risk-reducing intermediary among customers engaged in production. For example, the insurance industry pools premiums of many customers and must pay off only the few who actually incur losses. But risk arises in situations for which pooled-premium insurance is unavailable. For instance, as a hedge against higher fuel costs, an airline may want to buy a security whose value will rise if oil prices rise. But who wants to sell such a security? The role of a financial institution is to design such a security, determine a "fair" price for it, and sell it to airlines. The security thus sold is usually "derivative" (i.e., its value is based on the value of other, identified securities). "Fair" in this context means that the financial institution earns just enough from selling the security to enable it to trade in other securities whose relation with oil prices is such that, if oil prices do indeed rise, the firm can pay off its increased obligation to the airlines. An "efficient" market is one in which risk-hedging securities are widely available at "fair" prices.

The Black-Scholes option pricing formula provided, for the first time, a theoretical method of fairly pricing a risk-hedging security. If an investment bank offers a derivative security at a price that is higher than "fair," it may be underbid. If it offers the security at less than the "fair" price, it runs the risk of substantial loss. This makes the bank reluctant to offer many of the derivative securities that would contribute to market efficiency. In particular, the bank only wants to offer derivative securities whose "fair" price can be determined in advance. Furthermore, if the bank sells such a security, it must then address the hedging problem: how should it manage the risk associated with its new position? The mathematical theory growing out of the Black-Scholes option pricing formula provides solutions for both the pricing and hedging problems. It thus has enabled the creation of a host of specialized derivative securities. This theory is the subject of this text.

Relationship between Volumes I and II

Volume II treats the continuous-time theory of stochastic calculus within the context of finance applications. The presentation of this theory is the raison d'être of this work. Volume II includes a self-contained treatment of the probability theory needed for stochastic calculus, including Brownian motion and its properties.

Volume I presents many of the same finance applications, but within the simpler context of the discrete-time binomial model. It prepares the reader for Volume II by treating several fundamental concepts, including martingales, Markov processes, change of measure and risk-neutral pricing in this less technical setting. However, Volume II has a self-contained treatment of these topics, and strictly speaking, it is not necessary to read Volume I before reading Volume II. It is helpful in that the difficult concepts of Volume II are first seen in a simpler context in Volume I.

In the Carnegie Mellon Master's program in Computational Finance, the course based on Volume I is a prerequisite for the courses based on Volume II. However, graduate students in computer science, finance, mathematics, physics and statistics frequently take the courses based on Volume II without first taking the course based on Volume I.

The reader who begins with Volume II may use Volume I as a reference. As several concepts are presented in Volume II, reference is made to the analogous concepts in Volume I. The reader can at that point choose to read only Volume II or to refer to Volume I for a discussion of the concept at hand in a more transparent setting.

Summary of Volume I

Volume I presents the binomial asset pricing model. Although this model is interesting in its own right, and is often the paradigm of practice, here it is used primarily as a vehicle for introducing in a simple setting the concepts needed for the continuous-time theory of Volume II.

Chapter 1, *The Binomial No-Arbitrage Pricing Model*, presents the no-arbitrage method of option pricing in a binomial model. The mathematics is simple, but the profound concept of risk-neutral pricing introduced here is not. Chapter 2, *Probability Theory on Coin Toss Space*, formalizes the results of Chapter 1, using the notions of martingales and Markov processes. This chapter culminates with the risk-neutral pricing formula for European derivative securities. The tools used to derive this formula are not really required for the derivation in the binomial model, but we need these concepts in Volume II and therefore develop them in the simpler discrete-time setting of Volume I. Chapter 3, *State Prices*, discusses the change of measure associated with risk-neutral pricing of European derivative securities, again as a warm-up exercise for change of measure in continuous-time models. An interesting application developed here is to solve the problem of optimal (in the sense of expected utility maximization) investment in a binomial model. The ideas of Chapters 1 to 3 are essential to understanding the methodology of modern quantitative finance. They are developed again in Chapters 4 and 5 of Volume II.

The remaining three chapters of Volume I treat more specialized concepts. Chapter 4, *American Derivative Securities*, considers derivative securities whose owner can choose the exercise time. This topic is revisited in a continuous-time context in Chapter 8 of Volume II. Chapter 5, *Random Walk*, explains the reflection principle for random walk. The analogous reflection principle for Brownian motion plays a prominent role in the derivation of pricing formulas for exotic options in Chapter 7 of Volume II. Finally, Chapter 6, *Interest-Rate-Dependent Assets*, considers models with random interest rates, examining the difference between forward and futures prices and introducing the concept of a forward measure. Forward and futures prices reappear at the end of Chapter 5 of Volume II. Forward measures for continuous-time models are developed in Chapter 9 of Volume II and used to create forward LIBOR models for interest rate movements in Chapter 10 of Volume II.

Summary of Volume II

Chapter 1, *General Probability Theory*, and Chapter 2, *Information and Conditioning*, of Volume II lay the measure-theoretic foundation for probability theory required for a treatment of continuous-time models. Chapter 1 presents probability spaces, Lebesgue integrals, and change of measure. Independence, conditional expectations, and properties of conditional expectations are introduced in Chapter 2. These chapters are used extensively throughout the text, but some readers, especially those with exposure to probability theory, may choose to skip this material at the outset, referring to it as needed.

Chapter 3, *Brownian Motion*, introduces Brownian motion and its properties. The most important of these for stochastic calculus is quadratic variation, presented in Section 3.4. All of this material is needed in order to proceed, except Sections 3.6 and 3.7, which are used only in Chapter 7, *Exotic Options* and Chapter 8, *Early Exercise*.

The core of Volume II is Chapter 4, *Stochastic Calculus*. Here the Itô integral is constructed and Itô's formula (called the Itô-Doeblin formula in this text) is developed. Several consequences of the Itô-Doeblin formula are worked out. One of these is the characterization of Brownian motion in terms of its quadratic variation (Lévy's theorem) and another is the Black-Scholes equation for a European call price (called the Black-Scholes-Merton equation in this text). The only material which the reader may omit is Section 4.7, *Brownian Bridge*. This topic is included because of its importance in Monte Carlo simulation, but it is not used elsewhere in the text.

Chapter 5, *Risk-Neutral Pricing*, states and proves Girsanov's Theorem, which underlies change of measure. This permits a systematic treatment of risk-neutral pricing and the Fundamental Theorems of Asset Pricing (Section 5.4). Section 5.5, *Dividend-Paying Stocks*, is not used elsewhere in the text. Section 5.6, *Forwards and Futures*, appears later in Section 9.4 and in some exercises.

Chapter 6, *Connections with Partial Differential Equations*, develops the connection between stochastic calculus and partial differential equations. This is used frequently in later chapters.

With the exceptions noted above, the material in Chapters 1–6 is fundamental for quantitative finance is essential for reading the later chapters. After Chapter 6, the reader has choices.

Chapter 7, *Exotic Options*, is not used in subsequent chapters, nor is Chapter 8, *Early Exercise*. Chapter 9, *Change of Numéraire*, plays an important role in Section 10.4, *Forward LIBOR model*, but is not otherwise used. Chapter 10, *Term Structure Models*, and Chapter 11, *Introduction to Jump Processes*, are not used elsewhere in the text.

1

The Binomial No-Arbitrage Pricing Model

1.1 One-Period Binomial Model

The *binomial asset-pricing model* provides a powerful tool to understand *arbitrage pricing theory* and probability. In this chapter, we introduce this tool for the first purpose, and we take up the second in Chapter 2. In this section, we consider the simplest binomial model, the one with only one period. This is generalized to the more realistic multiperiod binomial model in the next section.

For the general one-period model of Figure 1.1.1, we call the beginning of the period *time zero* and the end of the period *time one*. At time zero, we have a stock whose price per share we denote by S_0, a positive quantity known at time zero. At time one, the price per share of this stock will be one of two positive values, which we denote $S_1(H)$ and $S_1(T)$, the H and T standing for *head* and *tail*, respectively. Thus, we are imagining that a coin is tossed, and the outcome of the coin toss determines the price at time one. We do not assume this coin is fair (i.e., the probability of head need not be one-half). We assume only that the probability of head, which we call p, is positive, and the probability of tail, which is $q = 1 - p$, is also positive.

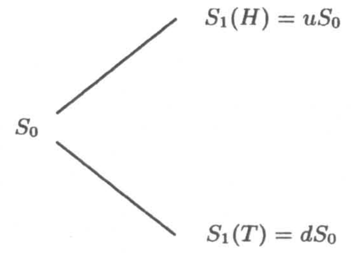

Fig. 1.1.1. General one-period binomial model.

The outcome of the coin toss, and hence the value which the stock price will take at time one, is known at time one but not at time zero. We shall refer to any quantity not known at time zero as *random* because it depends on the random experiment of tossing a coin.

We introduce the two positive numbers

$$u = \frac{S_1(H)}{S_0}, \quad d = \frac{S_1(T)}{S_0}. \tag{1.1.1}$$

We assume that $d < u$; if we instead had $d > u$, we may achieve $d < u$ by relabeling the sides of our coin. If $d = u$, the stock price at time one is not really random and the model is uninteresting. We refer to u as the *up factor* and d as the *down factor*. It is intuitively helpful to think of u as greater than one and to think of d as less than one, and hence the names *up factor* and *down factor*, but the mathematics we develop here does not require that these inequalities hold.

We introduce also an *interest rate* r. One dollar invested in the money market at time zero will yield $1+r$ dollars at time one. Conversely, one dollar borrowed from the money market at time zero will result in a debt of $1+r$ at time one. In particular, the interest rate for borrowing is the same as the interest rate for investing. It is almost always true that $r \geq 0$, and this is the case to keep in mind. However, the mathematics we develop requires only that $r > -1$.

An essential feature of an efficient market is that if a trading strategy can turn nothing into something, then it must also run the risk of loss. Otherwise, there would be an *arbitrage*. More specifically, we define *arbitrage* as a trading strategy that begins with no money, has zero probability of losing money, and has a positive probability of making money. A mathematical model that admits arbitrage cannot be used for analysis. Wealth can be generated from nothing in such a model, and the questions one would want the model to illuminate are provided with paradoxical answers by the model. Real markets sometimes exhibit arbitrage, but this is necessarily fleeting; as soon as someone discovers it, trading takes places that removes it.

In the one-period binomial model, to rule out arbitrage we must assume

$$0 < d < 1 + r < u. \tag{1.1.2}$$

The inequality $d > 0$ follows from the positivity of the stock prices and was already assumed. The two other inequalities in (1.1.2) follow from the absence of arbitrage, as we now explain. If $d \geq 1+r$, one could begin with zero wealth and at time zero borrow from the money market in order to buy stock. Even in the worst case of a tail on the coin toss, the stock at time one will be worth enough to pay off the money market debt and has a positive probability of being worth strictly more since $u > d \geq 1 + r$. This provides an arbitrage. On the other hand, if $u \leq 1 + r$, one could sell the stock short and invest the proceeds in the money market. Even in the best case for the stock, the cost of replacing it at time one will be less than or equal to the value of the money market investment, and since $d < u \leq 1 + r$, there is a positive probability that the cost of replacing the stock will be strictly less than the value of the money market investment. This again provides an arbitrage.

We have argued in the preceding paragraph that if there is to be no arbitrage in the market with the stock and the money market account, then we must have (1.1.2). The converse of this is also true. If (1.1.2) holds, then there is no arbitrage. See Exercise 1.1.

It is common to have $d = \frac{1}{u}$, and this will be the case in many of our examples. However, for the binomial asset-pricing model to make sense, we only need to assume (1.1.2).

Of course, stock price movements are much more complicated than indicated by the binomial asset-pricing model. We consider this simple model for three reasons. First of all, within this model, the concept of arbitrage pricing and its relation to risk-neutral pricing is clearly illuminated. Secondly, the model is used in practice because, with a sufficient number of periods, it provides a reasonably good, computationally tractable approximation to continuous-time models. Finally, within the binomial asset-pricing model, we can develop the theory of conditional expectations and martingales, which lies at the heart of continuous-time models.

Let us now consider a *European call option*, which confers on its owner the right but not the obligation to buy one share of the stock at time one for the *strike price* K. The interesting case, which we shall assume here, is that $S_1(T) < K < S_1(H)$. If we get a tail on the toss, the option expires worthless. If we get a head on the coin toss, the option can be *exercised* and yields a profit of $S_1(H) - K$. We summarize this situation by saying that the option at time one is worth $(S_1 - K)^+$, where the notation $(\cdots)^+$ indicates that we take the maximum of the expression in parentheses and zero. Here we follow the usual custom in probability of omitting the argument of the random variable S_1. The fundamental question of option pricing is how much the option is worth at time zero before we know whether the coin toss results in head or tail.

The *arbitrage pricing theory* approach to the option-pricing problem is to replicate the option by trading in the stock and money markets. We illustrate this with an example, and then we return to the general one-period binomial model.

Example 1.1.1. For the particular one-period model of Figure 1.1.2, let $S(0) = 4$, $u = 2$, $d = \frac{1}{2}$, and $r = \frac{1}{4}$. Then $S_1(H) = 8$ and $S_1(T) = 2$. Suppose the strike price of the European call option is $K = 5$. Suppose further that we begin with an initial wealth $X_0 = 1.20$ and buy $\Delta_0 = \frac{1}{2}$ shares of stock at time zero. Since stock costs 4 per share at time zero, we must use our initial wealth $X_0 = 1.20$ and borrow an additional 0.80 to do this. This leaves us with a cash position $X_0 - \Delta_0 S_0 = -0.80$ (i.e., a debt of 0.80 to the money market). At time one, our cash position will be $(1 + r)(X_0 - \Delta_0 S_0) = -1$

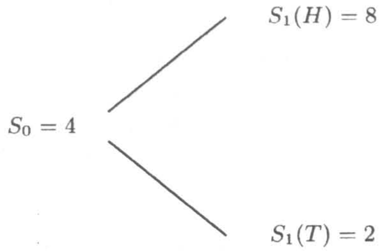

Fig. 1.1.2. Particular one-period binomial model.

(i.e., we will have a debt of 1 to the money market). On the other hand, at time one we will have stock valued at either $\frac{1}{2}S_1(H) = 4$ or $\frac{1}{2}S_1(T) = 1$. In particular, if the coin toss results in a head, the value of our portfolio of stock and money market account at time one will be

$$X_1(H) = \frac{1}{2}S_1(H) + (1+r)(X_0 - \Delta_0 S_0) = 3;$$

if the coin toss results in a tail, the value of our portfolio of stock and money market account at time one will be

$$X_1(T) = \frac{1}{2}S_1(T) + (1+r)(X_0 - \Delta_0 S_0) = 0.$$

In either case, the value of the portfolio agrees with the value of the option at time one, which is either $(S_1(H) - 5)^+ = 3$ or $(S_1(T) - 5)^+ = 0$. We have *replicated* the option by trading in the stock and money markets.

The initial wealth 1.20 needed to set up the replicating portfolio described above is the *no-arbitrage price of the option at time zero.* If one could sell the option for more than this, say, for 1.21, then the seller could invest the excess 0.01 in the money market and use the remaining 1.20 to replicate the option. At time one, the seller would be able to pay off the option, regardless of how the coin tossing turned out, and still have the 0.0125 resulting from the money market investment of the excess 0.01. This is an arbitrage because the seller of the option needs no money initially, and without risk of loss has 0.0125 at time one. On the other hand, if one could buy the option above for less than 1.20, say, for 1.19, then one should buy the option and set up the reverse of the replicating trading strategy described above. In particular, sell short one-half share of stock, which generates income 2. Use 1.19 to buy the option, put 0.80 in the money market, and in a separate money market account put the remaining 0.01. At time one, if there is a head, one needs 4 to replace the half-share of stock. The option bought at time zero is worth 3, and the 0.80 invested in the money market at time zero has grown to 1. At time one, if there is a tail, one needs 1 to replace the half-share of stock.

The option is worthless, but the 0.80 invested in the money market at time zero has grown to 1. In either case, the buyer of the option has a net zero position at time one, plus the separate money market account in which 0.01 was invested at time zero. Again, there is an arbitrage.

We have shown that in the market with the stock, the money market, and the option, there is an arbitrage unless the time-zero price of the option is 1.20. If the time-zero price of the option is 1.20, then there is no arbitrage (see Exercise 1.2). □

The argument in the example above depends on several assumptions. The principal ones are:

- shares of stock can be subdivided for sale or purchase,
- the interest rate for investing is the same as the interest rate for borrowing,
- the purchase price of stock is the same as the selling price (i.e., there is zero *bid–ask spread*),
- at any time, the stock can take only two possible values in the next period.

All these assumptions except the last also underlie the Black-Scholes-Merton option-pricing formula. The first of these assumptions is essentially satisfied in practice because option pricing and hedging (replication) typically involve lots of options. If we had considered 100 options rather than one option in Example 1.1.1, we would have hedged the short position by buying $\Delta_0 = 50$ shares of stock rather than $\Delta_0 = \frac{1}{2}$ of a share. The second assumption is close to being true for large institutions. The third assumption is not satisfied in practice. Sometimes the bid–ask spread can be ignored because not too much trading is taking place. In other situations, this departure of the model from reality becomes a serious issue. In the Black-Scholes-Merton model, the fourth assumption is replaced by the assumption that the stock price is a geometric Brownian motion. Empirical studies of stock price returns have consistently shown this not to be the case. Once again, the departure of the model from reality can be significant in some situations, but in other situations the model works remarkably well. We shall develop a modeling framework that extends far beyond the geometric Brownian motion assumption, a framework that includes many of the more sophisticated models that are not tied to this assumption.

In the general one-period model, we define a *derivative security* to be a security that pays some amount $V_1(H)$ at time one if the coin toss results in head and pays a possibly different amount $V_1(T)$ at time one if the coin toss results in tail. A European call option is a particular kind of derivative security. Another is the *European put option,* which pays off $(K - S_1)^+$ at time one, where K is a constant. A third is a *forward contract,* whose value at time one is $S_1 - K$.

To determine the price V_0 at time zero for a derivative security, we replicate it as in Example 1.1.1. Suppose we begin with wealth X_0 and buy Δ_0 shares of stock at time zero, leaving us with a cash position $X_0 - \Delta_0 S_0$. The value of our portfolio of stock and money market account at time one is

$$X_1 = \Delta_0 S_1 + (1+r)(X_0 - \Delta_0 S_0) = (1+r)X_0 + \Delta_0(S_1 - (1+r)S_0).$$

We want to choose X_0 and Δ_0 so that $X_1(H) = V_1(H)$ and $X_1(T) = V_1(T)$. (Note here that $V_1(H)$ and $V_1(T)$ are given quantities, the amounts the derivative security will pay off depending on the outcome of the coin tosses. At time zero, we know what the two possibilities $V_1(H)$ and $V_1(T)$ are; we do not know which of these two possibilities will be realized.) Replication of the derivative security thus requires that

$$X_0 + \Delta_0\left(\frac{1}{1+r}S_1(H) - S_0\right) = \frac{1}{1+r}V_1(H), \tag{1.1.3}$$

$$X_0 + \Delta_0\left(\frac{1}{1+r}S_1(T) - S_0\right) = \frac{1}{1+r}V_1(T). \tag{1.1.4}$$

One way to solve these two equations in two unknowns is to multiply the first by a number $\tilde{p}$ and the second by $\tilde{q} = 1 - \tilde{p}$ and then add them to get

$$X_0 + \Delta_0\left(\frac{1}{1+r}\left[\tilde{p}S_1(H) + \tilde{q}S_1(T)\right] - S_0\right) = \frac{1}{1+r}\left[\tilde{p}V_1(H) + \tilde{q}V_1(T)\right]. \tag{1.1.5}$$

If we choose $\tilde{p}$ so that

$$S_0 = \frac{1}{1+r}\left[\tilde{p}S_1(H) + \tilde{q}S_1(T)\right], \tag{1.1.6}$$

then the term multiplying Δ_0 in (1.1.5) is zero, and we have the simple formula for X_0

$$X_0 = \frac{1}{1+r}\left[\tilde{p}V_1(H) + \tilde{q}V_1(T)\right]. \tag{1.1.7}$$

We can solve for $\tilde{p}$ directly from (1.1.6) in the form

$$S_0 = \frac{1}{1+r}\left[\tilde{p}uS_0 + (1-\tilde{p})dS_0\right] = \frac{S_0}{1+r}\left[(u-d)\tilde{p} + d\right].$$

This leads to the formulas

$$\tilde{p} = \frac{1+r-d}{u-d}, \quad \tilde{q} = \frac{u-1-r}{u-d}. \tag{1.1.8}$$

We can solve for Δ_0 by simply subtracting (1.1.4) from (1.1.3) to get the *delta-hedging formula*

$$\Delta_0 = \frac{V_1(H) - V_1(T)}{S_1(H) - S_1(T)}. \tag{1.1.9}$$

In conclusion, if an agent begins with wealth X_0 given by (1.1.7) and at time zero buys Δ_0 shares of stock, given by (1.1.9), then at time one, if the coin toss results in head, the agent will have a portfolio worth $V_1(H)$, and if the coin toss results in tail, the portfolio will be worth $V_1(T)$. The agent has *hedged a short position in the derivative security*. The derivative security that pays V_1 at time one should be priced at

$$V_0 = \frac{1}{1+r}\left[\tilde{p}V_1(H) + \tilde{q}V_1(T)\right] \tag{1.1.10}$$

at time zero. This price permits the seller to hedge the short position in the claim. This price does not introduce an arbitrage when the derivative security is added to the market comprising the stock and money market account; any other time-zero price would introduce an arbitrage.

Although we have determined the no-arbitrage price of a derivative security by setting up a hedge for a short position in the security, one could just as well consider the hedge for a long position. An agent with a long position owns an asset having a certain value, and the agent may wish to set up a hedge to protect against loss of that value. This is how practitioners think about hedging. The number of shares of the underlying stock held by a long position hedge is the negative of the number determined by (1.1.9). Exercises 1.6 and 1.7 consider this is more detail.

The numbers $\tilde{p}$ and $\tilde{q}$ given by (1.1.8) are both positive because of the no-arbitrage condition (1.1.2), and they sum to one. For this reason, we can regard them as probabilities of head and tail, respectively. They are not the actual probabilities, which we call p and q, but rather the so-called *risk-neutral probabilities*. Under the actual probabilities, the average rate of growth of the stock is typically strictly greater than the rate of growth of an investment in the money market; otherwise, no one would want to incur the risk associated with investing in the stock. Thus, p and $q = 1 - p$ should satisfy

$$S_0 < \frac{1}{1+r}\left[pS_1(H) + qS_1(T)\right],$$

whereas $\tilde{p}$ and $\tilde{q}$ satisfy (1.1.6). If the average rate of growth of the stock were exactly the same as the rate of growth of the money market investment, then investors must be neutral about risk—they do not require compensation for assuming it, nor are they willing to pay extra for it. This is simply not the case, and hence $\tilde{p}$ and $\tilde{q}$ cannot be the actual probabilities. They are only numbers that assist us in the solution of the two equations (1.1.3) and (1.1.4) in the two unknowns X_0 and Δ_0. They assist us by making the term multiplying the unknown Δ_0 in (1.1.5) drop out. In fact, because they are chosen to make the mean rate of growth of the stock appear to equal the rate of growth of the money market account, they make the mean rate of growth of any portfolio of stock and money market account appear to equal the rate of growth of the money market asset. If we want to construct a portfolio whose value at time one is V_1, then its value at time zero must be given by (1.1.7), so that its mean rate of growth under the risk-neutral probabilities is the rate of growth of the money market investment.

The concluding equation (1.1.10) for the time-zero price V_0 of the derivative security V_1 is called the *risk-neutral pricing formula* for the one-period binomial model. One should not be concerned that the actual probabilities do not appear in this equation. We have constructed a hedge for a short position in the derivative security, and this hedge works regardless of whether the stock goes up or down. The probabilities of the up and down moves are irrelevant. What matters is the size of the two possible moves (the values of u and d). In the binomial model, the prices of derivative securities depend on the set of possible stock price paths but not on how probable these paths are. As we shall see in Chapters 4 and 5 of Volume II, the analogous fact for continuous-time models is that prices of derivative securities depend on the volatility of stock prices but not on their mean rates of growth.

1.2 Multiperiod Binomial Model

We now extend the ideas in Section 1.1 to multiple periods. We toss a coin repeatedly, and whenever we get a head the stock price moves "up" by the factor u, whereas whenever we get a tail, the stock price moves "down" by the factor d. In addition to this stock, there is a money market asset with a constant interest rate r. The only assumption we make on these parameters is the no-arbitrage condition (1.1.2).

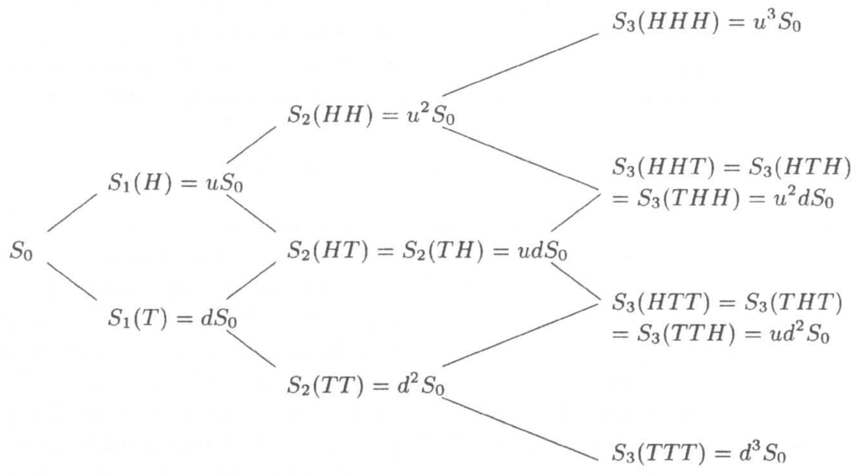

Fig. 1.2.1. General three-period model.

We denote the initial stock price by S_0, which is positive. We denote the price at time one by $S_1(H) = uS_0$ if the first toss results in head and by $S_1(T) = dS_0$ if the first toss results in tail. After the second toss, the price will be one of:

$$S_2(HH) = uS_1(H) = u^2 S_0,\ S_2(HT) = dS_1(H) = duS_0,$$
$$S_2(TH) = uS_1(T) = udS_0,\ S_2(TT) = dS_1(T) = d^2 S_0.$$

After three tosses, there are eight possible coin sequences, although not all of them result in different stock prices at time 3. See Figure 1.2.1.

Example 1.2.1. Consider the particular three-period model with $S_0 = 4$, $u = 2$, and $d = \frac{1}{2}$. We have the binomial "tree" of possible stock prices shown in Figure 1.2.2. □

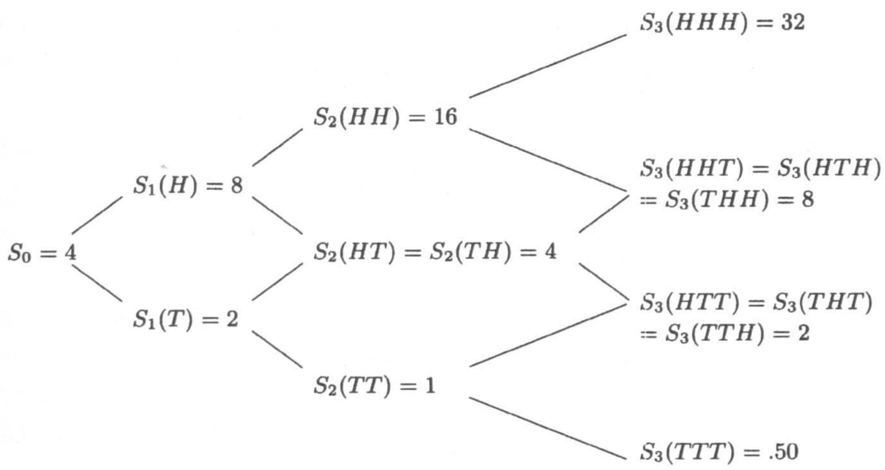

Fig. 1.2.2. A particular three-period model.

Let us return to the general three-period binomial model of Figure 1.2.1 and consider a European call that confers the right to buy one share of stock for K dollars at time two. After the discussion of this option, we extend the analysis to an arbitrary European derivative security that expires at time $N \geq 2$.

At expiration, the payoff of a call option with strike price K and expiration time two is $V_2 = (S_2 - K)^+$, where V_2 and S_2 depend on the first and second coin tosses. We want to determine the no-arbitrage price for this option at time zero. Suppose an agent sells the option at time zero for V_0 dollars, where V_0 is still to be determined. She then buys Δ_0 shares of stock, investing $V_0 - \Delta_0 S_0$ dollars in the money market to finance this. (The quantity $V_0 - \Delta_0 S_0$ will turn out to be negative, so the agent is actually borrowing $\Delta_0 S_0 - V_0$ dollars from the money market.) At time one, the agent has a portfolio (excluding the short position in the option) valued at

$$X_1 = \Delta_0 S_1 + (1+r)(V_0 - \Delta_0 S_0). \tag{1.2.1}$$

Although we do not indicate it in the notation, S_1 and therefore X_1 depend on the outcome of the first coin toss. Thus, there are really two equations implicit in (1.2.1):

$$X_1(H) = \Delta_0 S_1(H) + (1+r)(V_0 - \Delta_0 S_0), \tag{1.2.2}$$
$$X_1(T) = \Delta_0 S_1(T) + (1+r)(V_0 - \Delta_0 S_0). \tag{1.2.3}$$

After the first coin toss, the agent has a portfolio valued at X_1 dollars and can readjust her hedge. Suppose she decides now to hold Δ_1 shares of stock, where Δ_1 is allowed to depend on the first coin toss because the agent knows the result of this toss at time one when she chooses Δ_1. She invests the remainder of her wealth, $X_1 - \Delta_1 S_1$, in the money market. In the next period, her wealth will be given by the right-hand side of the following equation, and she wants it to be V_2. Therefore, she wants to have

$$V_2 = \Delta_1 S_2 + (1+r)(X_1 - \Delta_1 S_1). \tag{1.2.4}$$

Although we do not indicate it in the notation, S_2 and V_2 depend on the outcomes of the first two coin tosses. Considering all four possible outcomes, we can write (1.2.4) as four equations:

$$V_2(HH) = \Delta_1(H)S_2(HH) + (1+r)(X_1(H) - \Delta_1(H)S_1(H)), \tag{1.2.5}$$
$$V_2(HT) = \Delta_1(H)S_2(HT) + (1+r)(X_1(H) - \Delta_1(H)S_1(H)), \tag{1.2.6}$$
$$V_2(TH) = \Delta_1(T)S_2(TH) + (1+r)(X_1(T) - \Delta_1(T)S_1(T)), \tag{1.2.7}$$
$$V_2(TT) = \Delta_1(T)S_2(TT) + (1+r)(X_1(T) - \Delta_1(T)S_1(T)). \tag{1.2.8}$$

We now have six equations, the two represented by (1.2.1) and the four represented by (1.2.4), in the six unknowns V_0, Δ_0, $\Delta_1(H)$, $\Delta_1(T)$, $X_1(H)$, and $X_1(T)$.

To solve these equations, and thereby determine the no-arbitrage price V_0 at time zero of the option and the replicating portfolio Δ_0, $\Delta_1(H)$, and $\Delta_1(T)$, we begin with the last two equations, (1.2.7) and (1.2.8). Subtracting (1.2.8) from (1.2.7) and solving for $\Delta_1(T)$, we obtain the *delta-hedging formula*

$$\Delta_1(T) = \frac{V_2(TH) - V_2(TT)}{S_2(TH) - S_2(TT)}, \tag{1.2.9}$$

and substituting this into either (1.2.7) or (1.2.8), we can solve for

$$X_1(T) = \frac{1}{1+r}[\tilde{p}V_2(TH) + \tilde{q}V_2(TT)], \tag{1.2.10}$$

where $\tilde{p}$ and $\tilde{q}$ are the risk-neutral probabilities given by (1.1.8). We can also obtain (1.2.10) by multiplying (1.2.7) by $\tilde{p}$ and (1.2.8) by $\tilde{q}$ and adding them together. Since

$$\tilde{p}S_2(TH) + \tilde{q}S_2(TT) = (1+r)S_1(T),$$

this causes all the terms involving $\Delta_1(T)$ to drop out. Equation (1.2.10) gives the value the replicating portfolio should have at time one if the stock goes down between times zero and one. We define this quantity to be the *price of the option at time one if the first coin toss results in tail*, and we denote it by $V_1(T)$. We have just shown that

$$V_1(T) = \frac{1}{1+r}[\tilde{p}V_2(TH) + \tilde{q}V_2(TT)], \tag{1.2.11}$$

which is another instance of the *risk-neutral pricing formula*. This formula is analogous to formula (1.1.10) but postponed by one period. The first two equations, (1.2.5) and (1.2.6), lead in a similar way to the formulas

$$\Delta_1(H) = \frac{V_2(HH) - V_2(HT)}{S_2(HH) - S_2(HT)} \tag{1.2.12}$$

and $X_1(H) = V_1(H)$, where $V_1(H)$ is the *price of the option at time one if the first toss results in head*, defined by

$$V_1(H) = \frac{1}{1+r}[\tilde{p}V_2(HH) + \tilde{q}V_2(HT)]. \tag{1.2.13}$$

This is again analogous to formula (1.1.10), postponed by one period. Finally, we plug the values $X_1(H) = V_1(H)$ and $X_1(T) = V_1(T)$ into the two equations implicit in (1.2.1). The solution of these equations for Δ_0 and V_0 is the same as the solution of (1.1.3) and (1.1.4) and results again in (1.1.9) and (1.1.10).

To recap, we have three *stochastic processes*, (Δ_0, Δ_1), (X_0, X_1, X_2), and (V_0, V_1, V_2). By *stochastic process*, we mean a sequence of random variables indexed by time. These quantities are random because they depend on the coin tosses; indeed, the subscript on each variable indicates the number of coin tosses on which it depends. If we begin with any initial wealth X_0 and specify values for Δ_0, $\Delta_1(H)$, and $\Delta_1(T)$, then we can compute the value of the portfolio that holds the number of shares of stock indicated by these specifications and finances this by borrowing or investing in the money market as necessary. Indeed, the value of this portfolio is defined recursively, beginning with X_0, via the *wealth equation*

$$X_{n+1} = \Delta_n S_{n+1} + (1+r)(X_n - \Delta_n S_n). \tag{1.2.14}$$

One might regard this as a contingent equation; it defines *random variables*, and actual values of these random variables are not resolved until the outcomes of the coin tossing are revealed. Nonetheless, already at time zero this equation permits us to compute what the value of the portfolio will be at every subsequent time under every coin-toss scenario.

For a derivative security expiring at time two, the random variable V_2 is contractually specified in a way that is contingent upon the coin tossing (e.g., if the coin tossing results in $\omega_1\omega_2$, so the stock price at time two is $S_2(\omega_1\omega_2)$, then for the European call we have $V_2(\omega_1\omega_2) = (S_2(\omega_1\omega_2) - K)^+)$. We want to determine a value of X_0 and values for Δ_0, $\Delta_1(H)$, and $\Delta_1(T)$ so that X_2 given by applying (1.2.14) recursively satisfies $X_2(\omega_1\omega_2) = V_2(\omega_1\omega_2)$, regardless of the values of ω_1 and ω_2. The formulas above tell us how to do this. We call V_0 the value of X_0 that allows us to accomplish this, and we define $V_1(H)$ and $V_1(T)$ to be the values of $X_1(H)$ and $X_1(T)$ given by (1.2.14) when X_0 and Δ_0 are chosen by the prescriptions above. In general, we use the symbols Δ_n and X_n to represent the number of shares of stock held by the portfolio and the corresponding portfolio values, respectively, regardless of how the initial wealth X_0 and the Δ_n are chosen. When X_0 and the Δ_n are chosen to replicate a derivative security, we use the symbol V_n in place of X_n and call this the *(no-arbitrage) price of the derivative security at time n.*

The pattern that emerged with the European call expiring at time two persists, regardless of the number of periods and the definition of the final payoff of the derivative security. (At this point, however, we are considering only payoffs that come at a specified time; there is no possibility of early exercise.)

Theorem 1.2.2 (Replication in the multiperiod binomial model). *Consider an N-period binomial asset-pricing model, with $0 < d < 1 + r < u$, and with*

$$\tilde{p} = \frac{1+r-d}{u-d}, \quad \tilde{q} = \frac{u-1-r}{u-d}. \tag{1.2.15}$$

Let V_N be a random variable (a derivative security paying off at time N) depending on the first N coin tosses $\omega_1\omega_2\ldots\omega_N$. Define recursively backward in time the sequence of random variables $V_{N-1}, V_{N-2}, \ldots, V_0$ by

$$V_n(\omega_1\omega_2\ldots\omega_n) = \frac{1}{1+r}[\tilde{p}V_{n+1}(\omega_1\omega_2\ldots\omega_n H) + \tilde{q}V_{n+1}(\omega_1\omega_2\ldots\omega_n T)], \tag{1.2.16}$$

so that each V_n depends on the first n coin tosses $\omega_1\omega_2\ldots\omega_n$, where n ranges between $N-1$ and 0. Next define

$$\Delta_n(\omega_1\ldots\omega_n) = \frac{V_{n+1}(\omega_1\ldots\omega_n H) - V_{n+1}(\omega_1\ldots\omega_n T)}{S_{n+1}(\omega_1\ldots\omega_n H) - S_{n+1}(\omega_1\ldots\omega_n T)}, \tag{1.2.17}$$

where again n ranges between 0 and $N-1$. If we set $X_0 = V_0$ and define recursively forward in time the portfolio values $X_1, X_2, \ldots, X_N$ by (1.2.14), then we will have

$$X_N(\omega_1\omega_2\ldots\omega_N) = V_N(\omega_1\omega_2\ldots\omega_n) \text{ for all } \omega_1\omega_2\ldots\omega_N. \tag{1.2.18}$$

Definition 1.2.3. *For $n = 1, 2, \ldots, N$, the random variable $V_n(\omega_1\ldots\omega_n)$ in Theorem 1.2.2 is defined to be the* price of the derivative security at time n *if the outcomes of the first n tosses are $\omega_1\ldots\omega_n$. The* price of the derivative security at time zero *is defined to be V_0.*

PROOF OF THEOREM 1.2.2: We prove by forward induction on n that

$$X_n(\omega_1\omega_2\ldots\omega_n) = V_n(\omega_1\omega_2\ldots\omega_n) \text{ for all } \omega_1\omega_2\ldots\omega_n, \tag{1.2.19}$$

where n ranges between 0 and N. The case of $n = 0$ is given by the definition of X_0 as V_0. The case of $n = N$ is what we want to show.

For the induction step, we assume that (1.2.19) holds for some value of n less than N and show that it holds for $n+1$. We thus let $\omega_1\omega_2\ldots\omega_n\omega_{n+1}$ be fixed but arbitrary and assume as the induction hypothesis that (1.2.19) holds for the particular $\omega_1\omega_2\ldots\omega_n$ we have fixed. We don't know whether $\omega_{n+1} = H$ or $\omega_{n+1} = T$, so we consider both cases. We first use (1.2.14) to compute $X_{n+1}(\omega_1\omega_2\ldots\omega_n H)$, to wit

$$\begin{aligned} &X_{n+1}(\omega_1\omega_2\ldots\omega_n H) \\ &= \Delta_n(\omega_1\omega_2\ldots\omega_n)uS_n(\omega_1\omega_2\ldots\omega_n) \\ &\qquad +(1+r)\Big(X_n(\omega_1\omega_2\ldots\omega_n) - \Delta_n(\omega_1\omega_2\ldots\omega_n)S_n(\omega_1\omega_2\ldots\omega_n)\Big). \end{aligned}$$

To simplify the notation, we suppress $\omega_1\omega_2\ldots\omega_n$ and write this equation simply as

$$X_{n+1}(H) = \Delta_n uS_n + (1+r)(X_n - \Delta_n S_n). \tag{1.2.20}$$

With $\omega_1\omega_2\ldots\omega_n$ similarly suppressed, we have from (1.2.17) that

$$\Delta_n = \frac{V_{n+1}(H) - V_{n+1}(T)}{S_{n+1}(H) - S_{n+1}(T)} = \frac{V_{n+1}(H) - V_{n+1}(T)}{(u-d)S_n}.$$

Substituting this into (1.2.20) and using the induction hypothesis (1.2.19) and the definition (1.2.16) of V_n, we see that

$$\begin{aligned} X_{n+1}(H) &= (1+r)X_n + \Delta_n S_n(u - (1+r)) \\ &= (1+r)V_n + \frac{(V_{n+1}(H) - V_{n+1}(T))(u-(1+r))}{u-d} \\ &= (1+r)V_n + \tilde{q}V_{n+1}(H) - \tilde{q}V_{n+1}(T) \\ &= \tilde{p}V_{n+1}(H) + \tilde{q}V_{n+1}(T) + \tilde{q}V_{n+1}(H) - \tilde{q}V_{n+1}(T) \\ &= V_{n+1}(H). \end{aligned}$$

Reinstating the suppressed coin tosses $\omega_1\omega_2\ldots\omega_n$, we may write this as

$$X_{n+1}(\omega_1\omega_2\ldots\omega_n H) = V_{n+1}(\omega_1\omega_2\ldots\omega_n H).$$

A similar argument (see Exercise 1.4) shows that

$$X_{n+1}(\omega_1\omega_2\ldots\omega_n T) = V_{n+1}(\omega_1\omega_2\ldots\omega_n T).$$

Consequently, regardless of whether $\omega_{n+1} = H$ or $\omega_{n+1} = T$, we have

$$X_{n+1}(\omega_1\omega_2\ldots\omega_n\omega_{n+1}) = V_{n+1}(\omega_1\omega_2\ldots\omega_n\omega_{n+1}).$$

Since $\omega_1\omega_2\dots\omega_n\omega_{n+1}$ is arbitrary, the induction step is complete. □

The multiperiod binomial model of this section is said to be *complete* because every derivative security can be replicated by trading in the underlying stock and the money market. In a complete market, every derivative security has a unique price that precludes arbitrage, and this is the price of Definition 1.2.3.

Theorem 1.2.2 applies to so-called *path-dependent* options as well as to derivative securities whose payoff depends only on the final stock price. We illustrate this point with the following example.

Example 1.2.4. Suppose as in Figure 1.2.2 that $S_0 = 4$, $u = 2$, and $d = \frac{1}{2}$. Assume the interest rate is $r = \frac{1}{4}$. Then $\tilde{p} = \tilde{q} = \frac{1}{2}$. Consider a *lookback option* that pays off

$$V_3 = \max_{0 \le n \le 3} S_n - S_3$$

at time three. Then

$$\begin{aligned}
V_3(HHH) &= S_3(HHH) - S_3(HHH) &&= 32 - 32 = 0,\\
V_3(HHT) &= S_2(HH) - S_3(HHT) &&= 16 - 8 = 8,\\
V_3(HTH) &= S_1(H) - S_3(HTH) &&= 8 - 8 = 0,\\
V_3(HTT) &= S_1(H) - S_3(HTT) &&= 8 - 2 = 6,\\
V_3(THH) &= S_3(THH) - S_3(THH) &&= 8 - 8 = 0,\\
V_3(THT) &= S_2(TH) - S_3(THT) &&= 4 - 2 = 2,\\
V_3(TTH) &= S_0 - S_3(TTH) &&= 4 - 2 = 2,\\
V_3(TTT) &= S_0 - S_3(TTT) &&= 4 - 0.50 = 3.50.
\end{aligned}$$

We compute the price of the option at other times using the backward recursion (1.2.16). This gives

$$V_2(HH) = \frac{4}{5}\left[\frac{1}{2}V_3(HHH) + \frac{1}{2}V_3(HHT)\right] = 3.20,$$

$$V_2(HT) = \frac{4}{5}\left[\frac{1}{2}V_3(HTH) + \frac{1}{2}V_3(HTT)\right] = 2.40,$$

$$V_2(TH) = \frac{4}{5}\left[\frac{1}{2}V_3(THH) + \frac{1}{2}V_3(THT)\right] = 0.80,$$

$$V_2(TT) = \frac{4}{5}\left[\frac{1}{2}V_3(TTH) + \frac{1}{2}V_3(TTT)\right] = 2.20,$$

and then

$$V_1(H) = \frac{4}{5}\left[\frac{1}{2}V_2(HH) + \frac{1}{2}V_2(HT)\right] = 2.24,$$

$$V_1(T) = \frac{4}{5}\left[\frac{1}{2}V_2(TH) + \frac{1}{2}V_2(TT)\right] = 1.20,$$

and finally

$$V_0 = \frac{4}{5}\left[\frac{1}{2}V_1(H) + \frac{1}{2}V_1(T)\right] = 1.376.$$

If an agent sells the lookback option at time zero for 1.376, she can hedge her short position in the option by buying

$$\Delta_0 = \frac{V_1(H) - V_1(T)}{S_1(H) - S_1(T)} = \frac{2.24 - 1.20}{8 - 2} = 0.1733$$

shares of stock. This costs 0.6933 dollars, which leaves her with $1.376 - 0.6933 = 0.6827$ to invest in the money market at 25% interest. At time one, she will have 0.8533 in the money market. If the stock goes up in price to 8, then at time one her stock is worth 1.3867, and so her total portfolio value is 2.24, which is $V_1(H)$. If the stock goes down in price to 2, then at time one her stock is worth 0.3467 and so her total portfolio value is 1.20, which is $V_1(T)$. Continuing this process, the agent can be sure to have a portfolio worth V_3 at time three, no matter how the coin tossing turns out. □

1.3 Computational Considerations

The amount of computation required by a naive implementation of the derivative security pricing algorithm given in Theorem 1.2.2 grows exponentially with the number of periods. The binomial models used in practice often have 100 or more periods, and there are $2^{100} \approx 10^{30}$ possible outcomes for a sequence of 100 coin tosses. An algorithm that begins by tabulating 2^{100} values for V_{100} is not computationally practical.

Fortunately, the algorithm given in Theorem 1.2.2 can usually be organized in a computationally efficient manner. We illustrate this with two examples.

Example 1.3.1. In the model with $S_0 = 4$, $u = 2$, $d = \frac{1}{2}$ and $r = \frac{1}{4}$, consider the problem of pricing a European put with strike price $K = 5$, expiring at time three. The risk-neutral probabilities are $\tilde{p} = \frac{1}{2}$, $\tilde{q} = \frac{1}{2}$. The stock process is shown in Figure 1.2.2. The payoff of the option, given by $V_3 = (5 - S_3)^+$, can be tabulated as

$$\begin{gathered}
V_3(HHH) = 0, \quad V_3(HHT) = V_3(HTH) = V_3(THH) = 0\\
V_3(HTT) = V_3(THT) = V_3(TTH) = 3, \quad V_3(TTT) = 4.50.
\end{gathered}$$

There are $2^3 = 8$ entries in this table, but an obvious simplification is possible. Let us denote by $v_3(s)$ the payoff of the option at time three when the stock price at time three is s. Whereas V_3 has the sequence of three coin tosses as its argument, the argument of v_3 is a stock price. At time three there are only four possible stock prices, and we can tabulate the relevant values of v_3 as

$$v_3(32) = 0, \ v_3(8) = 0, \ v_3(2) = 3, \ v_3(.50) = 4.50.$$

If the put expired after 100 periods, the argument of V_{100} would range over the 2^{100} possible outcomes of the coin tosses whereas the argument of v_{100} would range over the 101 possible stock prices at time 100. This is a tremendous reduction in computational complexity.

According to Theorem 1.2.2, we compute V_2 by the formula

$$V_2(\omega_1\omega_2) = \frac{2}{5}\Big[V_3(\omega_1\omega_2 H) + V_3(\omega_1\omega_2 T)\Big]. \tag{1.3.1}$$

Equation (1.3.1) represents four equations, one for each possible choice of $\omega_1\omega_2$. We let $v_2(s)$ denote the price of the put at time two if the stock price at time two is s. In terms of this function, (1.3.1) takes the form

$$v_2(s) = \frac{2}{5}\Big[v_3(2s) + v_3(\tfrac{1}{2}s)\Big],$$

and this represents only three equations, one for each possible value of the stock price at time two. Indeed, we may compute

$$\begin{aligned}
v_2(16) &= \frac{2}{5}\Big[v_3(32) + v_3(8)\Big] = 0,\\
v_2(4) &= \frac{2}{5}\Big[v_3(8) + v_3(2)\Big] = 1.20,\\
v_2(1) &= \frac{2}{5}\Big[v_3(2) + v_3(.50)\Big] = 3.
\end{aligned}$$

Similarly,

$$\begin{aligned}
v_1(8) &= \frac{2}{5}\Big[v_2(16) + v_2(4)\Big] = 0.48,\\
v_1(2) &= \frac{2}{5}\Big[v_2(4) + v_2(1)\Big] = 1.68,
\end{aligned}$$

where $v_1(s)$ denotes the price of the put at time one if the stock price at time one is s. The price of the put at time zero is

$$v_0(4) = \frac{2}{5}\Big[v_1(8) + v_1(2)\Big] = 0.864.$$

At each time $n = 0, 1, 2$, if the stock price is s, the number of shares of stock that should be held by the replicating portfolio is

$$\delta_n(s) = \frac{v_{n+1}(2s) - v_{n+1}(\frac{1}{2}s)}{2s - \frac{1}{2}s}.$$

This is the analogue of formula (1.2.17). □

In Example 1.3.1, the price of the option at any time n was a function of the stock price S_n at that time and did not otherwise depend on the coin tosses. This permitted the introduction of the functions v_n related to the random variables V_n by the formula $V_n = v_n(S_n)$. A similar reduction is often possible when the price of the option does depend on the stock price path rather than just the current stock price. We illustrate this with a second example.

Example 1.3.2. Consider the lookback option of Example 1.2.4. At each time n, the price of the option can be written as a function of the stock price S_n and the maximum stock price $M_n = \max_{0\le k\le n} S_k$ to date. At time three, there are six possible pairs of values for (S_3, M_3), namely

$$(32,32),\ (8,16),\ (8,8),\ (2,8),\ (2,4),\ (.50,4).$$

We define $v_3(s,m)$ to be the payoff of the option at time three if $S_3 = s$ and $M_3 = m$. We have

$$\begin{aligned}
&v_3(32,32) = 0,\ v_3(8,16) = 8,\ v_3(8,8) = 0,\\
&v_3(2,8) = 6,\ v_3(2,4) = 2,\ v_3(.50,4) = 3.50.
\end{aligned}$$

In general, let $v_n(s,m)$ denote the value of the option at time n if $S_n = s$ and $M_n = m$. The algorithm of Theorem 1.2.2 can be rewritten in terms of the functions v_n as

$$v_n(s,m) = \frac{2}{5}\Big[v_{n+1}(2s, m \vee (2s)) + v_{n+1}(\tfrac{1}{2}s, m)\Big],$$

where $m \vee (2s)$ denotes the maximum of m and $2s$. Using this algorithm, we compute

$$\begin{aligned}
v_2(16,16) &= \frac{2}{5}\Big[v_3(32,32) + v_2(8,16)\Big] = 3.20,\\
v_2(4,8) &= \frac{2}{5}\Big[v_3(8,8) + v_3(2,8)\Big] = 2.40,\\
v_2(4,4) &= \frac{2}{5}\Big[v_3(8,8) + v_3(2,4)\Big] = 0.80,\\
v_2(1,4) &= \frac{2}{5}\Big[v_3(2,4) + v_3(.50,4)\Big] = 2.20,
\end{aligned}$$

then compute

$$\begin{aligned}
v_1(8,8) &= \frac{2}{5}\Big[v_2(16,16) + v_2(4,8)\Big] = 2.24,\\
v_1(2,4) &= \frac{2}{5}\Big[v_1(4,4) + v_1(1,4)\Big] = 1.20,
\end{aligned}$$

and finally obtain the time-zero price

$$v_0(4,4) = \frac{2}{5}\Big[v_1(8,8) + v_1(2,4)\Big] = 1.376.$$

At each time $n = 0, 1, 2$, if the stock price is s and the maximum stock price to date is m, the number of shares of stock that should be held by the replicating portfolio is

$$\delta_n(s, m) = \frac{v_{n+1}(2s, m \vee (2s)) - v_{n+1}(\frac{1}{2}s, m)}{2s - \frac{1}{2}s}.$$

This is the analogue of formula (1.2.17). □

1.4 Summary

This chapter considers a multiperiod binomial model. At each period in this model, we toss a coin whose outcome determines whether the stock price changes by a factor of u or a factor of d, where $0 < d < u$. In addition to the stock, there is a money market account with per-period rate of interest r. This is the rate of interest applied to both investing and borrowing.

Arbitrage is a trading strategy that begins with zero capital and trades in the stock and money markets in order to make money with positive probability without any possibility of losing money. The multiperiod binomial model admits no arbitrage if and only if

$$0 < d < 1 + r < u. \tag{1.1.2}$$

We shall always impose this condition.

A derivative security pays off at some expiration time N contingent upon the coin tosses in the first N periods. The *arbitrage pricing theory* method of assigning a price to a derivative security prior to expiration can be understood in two ways. First, one can ask how to assign a price so that one cannot form an arbitrage by trading in the derivative security, the underlying stock, and the money market. This no-arbitrage condition uniquely determines the price at all times of the derivative security. Secondly, at any time n prior to the expiration time N, one can imagine selling the derivative security for a price and using the income from this sale to form a portfolio, dynamically trading the stock and money market asset from time n until the expiration time N. This portfolio hedges the short position in the derivative security if its value at time N agrees with the payoff of the derivative security, regardless of the outcome of the coin tossing between times n and N. The amount for which the derivative security must be sold at time n in order to construct this hedge of the short position is the same no-arbitrage price obtained by the first pricing method.

The no-arbitrage price of the derivative security that pays V_N at time N can be computed recursively, backward in time, by the formula

$$V_n(\omega_1\omega_2\ldots\omega_n) = \frac{1}{1+r}[\tilde{p}V_{n+1}(\omega_1\omega_2\ldots\omega_n H) + \tilde{q}V_{n+1}(\omega_1\omega_2\ldots\omega_n T)]. \tag{1.2.16}$$

The number of shares of the stock that should be held by a portfolio hedging a short position in the derivative security is given by

$$\Delta_n(\omega_1\ldots\omega_n) = \frac{V_{n+1}(\omega_1\ldots\omega_n H) - V_{n+1}(\omega_1\ldots\omega_n T)}{S_{n+1}(\omega_1\ldots\omega_n H) - S_{n+1}(\omega_1\ldots\omega_n T)}. \tag{1.2.17}$$

The numbers $\tilde{p}$ and $\tilde{q}$ appearing in (1.2.16) are the *risk-neutral probabilities* given by

$$\tilde{p} = \frac{1+r-d}{u-d}, \quad \tilde{q} = \frac{u-1-r}{u-d}. \tag{1.2.15}$$

These risk-neutral probabilities are positive because of (1.1.2) and sum to 1. They have the property that, at any time, the price of the stock is the discounted risk-neutral average of its two possible prices at the next time:

$$S_n(\omega_1\ldots\omega_n) = \frac{1}{1+r}\big[\tilde{p}S_{n+1}(\omega_1\ldots\omega_n H) + \tilde{q}S_{n+1}(\omega_1\ldots\omega_n T)\big].$$

In other words, under the risk-neutral probabilities, the mean rate of return for the stock is r, the same as the rate of return for the money market. Therefore, if these probabilities actually governed the coin tossing (in fact, they do not), then an agent trading in the money market account and stock would have before him two opportunities, both of which provide the same mean rate of return. Consequently, no matter how he invests, the mean rate of return for his portfolio would also be r. In particular, if it is time $N-1$ and he wants his portfolio value to be $V_N(\omega_1\ldots\omega_N)$ at time N, then at time $N-1$ his portfolio value must be

$$\frac{1}{1+r}\big[\tilde{p}V_N(\omega_1\ldots\omega_{N-1}H) + \tilde{q}V_N(\omega_1\ldots\omega_{N-1}T)\big].$$

This is the right-hand side of (1.2.16) with $n = N-1$, and repeated application of this argument yields (1.2.16) for all values of n.

The explanation of (1.2.16) above was given under a condition contrary to fact, namely that $\tilde{p}$ and $\tilde{q}$ govern the coin tossing. One can ask whether such an argument can result in a valid conclusion. It does result in a valid conclusion for the following reason. When hedging a short position in a derivative security, we want the hedge to give us a portfolio that agrees with the payoff of the derivative security *regardless of the coin tossing.* In other words, the hedge must work *on all stock price paths.* If a path is possible (i.e., has positive probability), we want the hedge to work along that path. The actual value of the probability is irrelevant. We find these hedges by solving a system of equations along the paths, a system of the form (1.2.2)–(1.2.3), (1.2.5)–(1.2.8). There are no probabilities in this system. Introducing the risk-neutral probabilities allows us to argue as above and find a solution to the system. Introducing any other probabilities would not allow such an argument because only the risk-neutral probabilities allow us to state that no matter how the agent invests, the mean rate of return for his portfolio is r. The risk-neutral probabilities provide a shortcut to solving the system of equations. The actual probabilities are no help in solving this system. Under the actual probabilities, the mean rate of return for a portfolio depends on the portfolio, and when we are trying to solve the system of equations, we do not know which portfolio we should use.

Alternatively, one can explain (1.2.16) without recourse to any discussion of probability. This was the approach taken in the proof of Theorem 1.2.2. The numbers $\tilde{p}$ and $\tilde{q}$ were used in that proof, but they were not regarded as probabilities, just numbers defined by the formula (1.2.15).

1.5 Notes

No-arbitrage pricing is implicit in the work of Black and Scholes [5], but its first explicit development is provided by Merton [34], who began with the axiom of no-arbitrage and obtained a surprising number of conclusions. No arbitrage pricing was fully developed in continuous-time models by Harrison and Kreps [17] and Harrison and Pliska [18]. These authors introduced martingales (Sections 2.4 in this text and Section 2.3 in Volume II) and risk-neutral pricing. The binomial model is due to Cox, Ross, Rubinstein [11]; a good reference is [12]. The binomial model is useful in its own right, and as Cox et al. showed, one can rederive the Black-Scholes-Merton formula as a limit of the binomial model (see Theorem 3.2.2 in Chapter 3 of Volume II for the log-normality of the stock price obtained in the limit of the binomial model.)

1.6 Exercises

Exercise 1.1. Assume in the one-period binomial market of Section 1.1 that both H and T have positive probability of occurring. Show that condition (1.1.2) precludes arbitrage. In other words, show that if $X_0 = 0$ and

$$X_1 = \Delta_0 S_1 + (1+r)(X_0 - \Delta_0 S_0),$$

then we cannot have X_1 strictly positive with positive probability unless X_1 is strictly negative with positive probability as well, and this is the case regardless of the choice of the number Δ_0.

Exercise 1.2. Suppose in the situation of Example 1.1.1 that the option sells for 1.20 at time zero. Consider an agent who begins with wealth $X_0 = 0$ and at time zero buys Δ_0 shares of stock and Γ_0 options. The numbers Δ_0 and Γ_0 can be either positive or negative or zero. This leaves the agent with a cash position of $-4\Delta_0 - 1.20\Gamma_0$. If this is positive, it is invested in the money market; if it is negative, it represents money borrowed from the money market. At time one, the value of the agent's portfolio of stock, option, and money market assets is

$$X_1 = \Delta_0 S_1 + \Gamma_0 (S_1 - 5)^+ - \frac{5}{4}(4\Delta_0 + 1.20\Gamma_0).$$

Assume that both H and T have positive probability of occurring. Show that if there is a positive probability that X_1 is positive, then there is a positive probability that X_1 is negative. In other words, one cannot find an arbitrage when the time-zero price of the option is 1.20.

Exercise 1.3. In the one-period binomial model of Section 1.1, suppose we want to determine the price at time zero of the derivative security $V_1 = S_1$ (i.e., the derivative security pays off the stock price.) (This can be regarded as a European call with strike price $K = 0$). What is the time-zero price V_0 given by the risk-neutral pricing formula (1.1.10)?

Exercise 1.4. In the proof of Theorem 1.2.2, show under the induction hypothesis that

$$X_{n+1}(\omega_1\omega_2\ldots\omega_n T) = V_{n+1}(\omega_1\omega_2\ldots\omega_n T).$$

Exercise 1.5. In Example 1.2.4, we considered an agent who sold the lookback option for $V_0 = 1.376$ and bought $\Delta_0 = 0.1733$ shares of stock at time zero. At time one, if the stock goes up, she has a portfolio valued at $V_1(H) = 2.24$. Assume that she now takes a position of $\Delta_1(H) = \frac{V_2(HH)-V_2(HT)}{S_2(HH)-S_2(HT)}$ in the stock. Show that, at time two, if the stock goes up again, she will have a portfolio valued at $V_2(HH) = 3.20$, whereas if the stock goes down, her portfolio will be worth $V_2(HT) = 2.40$. Finally, under the assumption that the stock goes up in the first period and down in the second period, assume the agent takes a position of $\Delta_2(HT) = \frac{V_3(HTH)-V_3(HTT)}{S_3(HTH)-S_3(HTT)}$ in the stock. Show that, at time three, if the stock goes up in the third period, she will have a portfolio valued at $V_3(HTH) = 0$, whereas if the stock goes down, her portfolio will be worth $V_3(HTT) = 6$. In other words, she has hedged her short position in the option.

Exercise 1.6 (Hedging a long position-one period). Consider a bank that has a long position in the European call written on the stock price in Figure 1.1.2. The call expires at time one and has strike price $K = 5$. In Section 1.1, we determined the time-zero price of this call to be $V_0 = 1.20$. At time zero, the bank owns this option, which ties up capital $V_0 = 1.20$. The bank wants to earn the interest rate 25% on this capital until time one (i.e., without investing any more money, and regardless of how the coin tossing turns out, the bank wants to have

$$\frac{5}{4} \cdot 1.20 = 1.50$$

at time one, after collecting the payoff from the option (if any) at time one). Specify how the bank's trader should invest in the stock and money markets to accomplish this.

Exercise 1.7 (Hedging a long position-multiple periods). Consider a bank that has a long position in the lookback option of Example 1.2.4. The bank intends to hold this option until expiration and receive the payoff V_3. At time zero, the bank has capital $V_0 = 1.376$ tied up in the option and wants to earn the interest rate of 25% on this capital until time three (i.e., without investing any more money, and regardless of how the coin tossing turns out, the bank wants to have

$$\left(\frac{5}{4}\right)^3 \cdot 1.376 = 2.6875$$

at time three, after collecting the payoff from the lookback option at time three). Specify how the bank's trader should invest in the stock and the money market account to accomplish this.

Exercise 1.8 (Asian option). Consider the three-period model of Example 1.2.1, with $S_0 = 4$, $u = 2$, $d = \frac{1}{2}$, and take the interest rate $r = \frac{1}{4}$, so that $\tilde{p} = \tilde{q} = \frac{1}{2}$. For $n = 0, 1, 2, 3$, define $Y_n = \sum_{k=0}^{n} S_k$ to be the sum of the stock prices between times zero and n. Consider an *Asian call option* that expires at time three and has strike $K = 4$ (i.e., whose payoff at time three is $\left(\frac{1}{4}Y_3 - 4\right)^+$). This is like a European call, except the payoff of the option is based on the average stock price rather than the final stock price. Let $v_n(s, y)$ denote the price of this option at time n if $S_n = s$ and $Y_n = y$. In particular, $v_3(s, y) = \left(\frac{1}{4}y - 4\right)^+$.

(i) Develop an algorithm for computing v_n recursively. In particular, write a formula for v_n in terms of v_{n+1}.
(ii) Apply the algorithm developed in (i) to compute $v_0(4, 4)$, the price of the Asian option at time zero.
(iii) Provide a formula for $\delta_n(s, y)$, the number of shares of stock that should be held by the replicating portfolio at time n if $S_n = s$ and $Y_n = y$.

Exercise 1.9 (Stochastic volatility, random interest rate). Consider a binomial pricing model, but at each time $n \geq 1$, the "up factor" $u_n(\omega_1\omega_2 \ldots \omega_n)$, the "down factor" $d_n(\omega_1\omega_2 \ldots \omega_n)$, and the interest rate $r_n(\omega_1\omega_2 \ldots \omega_n)$ are allowed to depend on n and on the first n coin tosses $\omega_1\omega_2 \ldots \omega_n$. The initial up factor u_0, the initial down factor d_0, and the initial interest rate r_0 are not random. More specifically, the stock price at time one is given by

$$S_1(\omega_1) = \begin{cases} u_0 S_0 \text{ if } \omega_1 = H, \\ d_0 S_0 \text{ if } \omega_1 = T, \end{cases}$$

and, for $n \geq 1$, the stock price at time $n + 1$ is given by

$$S_{n+1}(\omega_1\omega_2 \ldots \omega_n\omega_{n+1}) = \begin{cases} u_n(\omega_1\omega_2 \ldots \omega_n)S_n(\omega_1\omega_2 \ldots \omega_n) \text{ if } \omega_{n+1} = H, \\ d_n(\omega_1\omega_2 \ldots \omega_n)S_n(\omega_1\omega_2 \ldots \omega_n) \text{ if } \omega_{n+1} = T. \end{cases}$$

One dollar invested in or borrowed from the money market at time zero grows to an investment or debt of $1 + r_0$ at time one, and, for $n \geq 1$, one dollar invested in or borrowed from the money market at time n grows to an investment or debt of $1 + r_n(\omega_1\omega_2 \ldots \omega_n)$ at time $n + 1$. We assume that for each n and for all $\omega_1\omega_2 \ldots \omega_n$, the no-arbitrage condition

$$0 < d_n(\omega_1\omega_2 \ldots \omega_n) < 1 + r_n(\omega_1\omega_2 \ldots \omega_n) < u_n(\omega_1\omega_2 \ldots \omega_n)$$

holds. We also assume that $0 < d_0 < 1 + r_0 < u_0$.

(i) Let N be a positive integer. In the model just described, provide an algorithm for determining the price at time zero for a derivative security that at time N pays off a random amount V_N depending on the result of the first N coin tosses.
(ii) Provide a formula for the number of shares of stock that should be held at each time n $(0 \leq n \leq N - 1)$ by a portfolio that replicates the derivative security V_N.
(iii) Suppose the initial stock price is $S_0 = 80$, with each head the stock price increases by 10, and with each tail the stock price decreases by 10. In other words, $S_1(H) = 90$, $S_1(T) = 70$, $S_2(HH) = 100$, etc. Assume the interest rate is always zero. Consider a European call with strike price 80, expiring at time five. What is the price of this call at time zero?

2

Probability Theory on Coin Toss Space

2.1 Finite Probability Spaces

A finite probability space is used to model a situation in which a random experiment with finitely many possible outcomes is conducted. In the context of the binomial model of the previous chapter, we tossed a coin a finite number of times. If, for example, we toss the coin three times, the set of all possible outcomes is

$$\Omega = \{HHH, HHT, HTH, HTT, THH, THT, TTH, TTT\}. \tag{2.1.1}$$

Suppose that on each toss the probability of a head (either actual or risk-neutral) is p and the probability of a tail is $q = 1-p$. We assume the tosses are independent, and so the probabilities of the individual elements ω (sequences of three tosses $\omega = \omega_1\omega_2\omega_3$) in Ω are

$$\begin{aligned}&\mathbb{P}(HHH) = p^3,\ \mathbb{P}(HHT) = p^2q,\ \mathbb{P}(HTH) = p^2q,\ \mathbb{P}(HTT) = pq^2,\\ &\mathbb{P}(THH) = p^2q,\ \mathbb{P}(THT) = pq^2,\ \mathbb{P}(TTH) = pq^2,\ \ \mathbb{P}(TTT) = q^3.\end{aligned} \tag{2.1.2}$$

The subsets of Ω are called *events*, and these can often be described in words as well as in symbols. For example, the event

$$\begin{aligned}\text{"The first toss is a head"} &= \{\omega \in \Omega; \omega_1 = H\}\\ &= \{HHH, HHT, HTH, HTT\}\end{aligned}$$

has, as indicated, descriptions in both words and symbols. We determine the probability of an event by summing the probabilities of the elements in the event, i.e.,

$$\begin{aligned}\mathbb{P}(\text{First toss is a head}) &= \mathbb{P}(HHH) + \mathbb{P}(HHT) + \mathbb{P}(HTH) + \mathbb{P}(HTT)\\ &= (p^3 + p^2q) + (p^2q + pq^2)\\ &= p^2(p+q) + pq(p+q)\end{aligned}$$

$$= p^2 + pq$$
$$= p(p+q)$$
$$= p. \tag{2.1.3}$$

Thus, the mathematics agrees with our intuition.

With mathematical models, it is easy to substitute our intuition for the mathematics, but this can lead to trouble. We should instead verify that the mathematics and our intuition agree; otherwise, either our intuition is wrong or our model is inadequate. If our intuition and the mathematics of a model do not agree, we should seek a reconciliation before proceeding. In the case of (2.1.3), we set out to build a model in which the probability of a head on each toss is p, we proposed doing this by defining the probabilities of the elements of Ω by (2.1.2), and we further defined the probability of an event (subset of Ω) to be the sum of the probabilities of the elements in the event. These definitions force us to carry out the computation (2.1.3) as we have done, and we need to do this computation in order to check that it gets the expected answer. Otherwise, we would have to rethink our mathematical model for the coin tossing.

We generalize slightly the situation just described, first by allowing Ω to be any finite set, and second by allowing some elements in Ω to have probability zero. These generalizations lead to the following definition.

Definition 2.1.1. *A* finite probability space *consists of a* sample space Ω *and a* probability measure $\mathbb{P}$. *The sample space Ω is a nonempty finite set and the probability measure $\mathbb{P}$ is a function that assigns to each element ω of Ω a number in $[0,1]$ so that*

$$\sum_{\omega \in \Omega} \mathbb{P}(\omega) = 1. \tag{2.1.4}$$

An event *is a subset of Ω, and we define the probability of an event A to be*

$$\mathbb{P}(A) = \sum_{\omega \in A} \mathbb{P}(\omega). \tag{2.1.5}$$

As mentioned before, this is a model for some random experiment. The set Ω is the set of all possible outcomes of the experiment, $\mathbb{P}(\omega)$ is the probability that the particular outcome ω occurs, and $\mathbb{P}(A)$ is the probability that the outcome that occurs is in the set A. If $\mathbb{P}(A) = 0$, then the outcome of the experiment is sure not to be in A; if $\mathbb{P}(A) = 1$, then the outcome is sure to be in A. Because of (2.1.4), we have the equation

$$\mathbb{P}(\Omega) = 1. \tag{2.1.6}$$

i.e., the outcome that occurs is sure to be in the set Ω. Because $\mathbb{P}(\omega)$ can be zero for some values of ω, we are permitted to put in Ω even some outcomes of the experiment that are sure not to occur. It is clear from (2.1.5) that if A and B are disjoint subsets of Ω, then

$$\mathbb{P}(A \cup B) = \mathbb{P}(A) + \mathbb{P}(B). \tag{2.1.7}$$

2.2 Random Variables, Distributions, and Expectations

A random experiment generally generates numerical data. This gives rise to the concept of a random variable.

Definition 2.2.1. *Let $(\Omega, \mathbb{P})$ be a finite probability space. A* random variable *is a real-valued function defined on Ω. (We sometimes also permit a random variable to take the values $+\infty$ and $-\infty$.)*

Example 2.2.2 (Stock prices). Recall the space Ω of three independent coin-tosses (2.1.1). As in Figure 1.2.2 of Chapter 1, let us define stock prices by the formulas

$$S_0(\omega_1\omega_2\omega_3) = 4 \text{ for all } \omega_1\omega_2\omega_3 \in \Omega_3,$$

$$S_1(\omega_1\omega_2\omega_3) = \begin{cases} 8 & \text{if } \omega_1 = H, \\ 2 & \text{if } \omega_1 = T, \end{cases}$$

$$S_2(\omega_1\omega_2\omega_3) = \begin{cases} 16 & \text{if } \omega_1 = \omega_2 = H, \\ 4 & \text{if } \omega_1 \neq \omega_2, \\ 1 & \text{if } \omega_1 = \omega_2 = T, \end{cases}$$

$$S_3(\omega_1\omega_2\omega_3) = \begin{cases} 32 & \text{if } \omega_1 = \omega_2 = \omega_3 = H, \\ 8 & \text{if there are two heads and one tail}, \\ 2 & \text{if there is one head and two tails}, \\ .50 & \text{if } \omega_1 = \omega_2 = \omega_3 = T. \end{cases}$$

Here we have written the arguments of S_0, S_1, S_2, and S_3 as $\omega_1\omega_2\omega_3$, even though some of these random variables do not depend on all the coin tosses. In particular, S_0 is actually not random because it takes the value 4, regardless of how the coin tosses turn out; such a random variable is sometimes called a *degenerate random variable.* □

It is customary to write the argument of random variables as ω, even when ω is a sequence such as $\omega = \omega_1\omega_2\omega_3$. We shall use these two notations interchangeably. It is even more common to write random variables without any arguments; we shall switch to that practice presently, writing S_3, for example, rather than $S_3(\omega_1\omega_2\omega_3)$ or $S_3(\omega)$.

According to Definition 2.2.1, a random variable is a function that maps a sample space Ω to the real numbers. The *distribution* of a random variable is a specification of the probabilities that the random variable takes various values. *A random variable is not a distribution*, and *a distribution is not a random variable.* This is an important point when we later switch between the actual probability measure, which one would estimate from historical data, and the risk-neutral probability measure. The change of measure will change distributions of random variables but not the random variables themselves. We make this distinction clear with the following example.

Example 2.2.3. Toss a coin three times, so the set of possible outcomes is

$$\Omega = \{HHH, HHT, HTH, HTT, THH, THT, TTH, TTT\}.$$

Define the random variables

$$X = \text{Total number of heads}, \qquad Y = \text{Total number of tails}.$$

In symbols,

$$\begin{aligned} X(HHH) &= 3,\\ X(HHT) = X(HTH) = X(THH) &= 2,\\ X(HTT) = X(THT) = X(TTH) &= 1,\\ X(TTT) &= 0, \end{aligned}$$

$$\begin{aligned} Y(TTT) &= 3,\\ Y(TTH) = Y(THT) = Y(HTT) &= 2,\\ Y(THH) = Y(HTH) = Y(HHT) &= 1,\\ Y(HHH) &= 0. \end{aligned}$$

We do not need to know probabilities of various outcomes in order to specify these random variables. However, once we specify a probability measure on Ω, we can determine the distributions of X and Y. For example, if we specify the probability measure $\widetilde{\mathbb{P}}$ under which the probability of head on each toss is $\frac{1}{2}$ and the probability of each element in Ω is $\frac{1}{8}$, then

$$\begin{aligned} \widetilde{\mathbb{P}}\{\omega \in \Omega; X(\omega) = 0\} &= \widetilde{\mathbb{P}}\{TTT\} = \frac{1}{8},\\ \widetilde{\mathbb{P}}\{\omega \in \Omega; X(\omega) = 1\} &= \widetilde{\mathbb{P}}\{HTT, THT, TTH\} = \frac{3}{8},\\ \widetilde{\mathbb{P}}\{\omega \in \Omega; X(\omega) = 2\} &= \widetilde{\mathbb{P}}\{HHT, HTH, THH\} = \frac{3}{8},\\ \widetilde{\mathbb{P}}\{\omega \in \Omega; X(\omega) = 3\} &= \widetilde{\mathbb{P}}\{HHH\} = \frac{1}{8}. \end{aligned}$$

We shorten the cumbersome notation $\widetilde{\mathbb{P}}\{\omega \in \Omega; X(\omega) = j\}$ to simply $\widetilde{\mathbb{P}}\{X = j\}$. It is helpful to remember, however, that the notation $\widetilde{\mathbb{P}}\{X = j\}$ refers to the probability of a subset of Ω, the set of elements ω for which $X(\omega) = j$. Under $\widetilde{\mathbb{P}}$, the probability that X takes the four values 0, 1, 2, and 3 are

$$\begin{aligned} \widetilde{\mathbb{P}}\{X = 0\} = \frac{1}{8}, \quad \widetilde{\mathbb{P}}\{X = 1\} = \frac{3}{8},\\ \widetilde{\mathbb{P}}\{X = 2\} = \frac{3}{8}, \quad \widetilde{\mathbb{P}}\{X = 3\} = \frac{1}{8}. \end{aligned}$$

This table of probabilities where X takes its various values records the *distribution* of X under $\widetilde{\mathbb{P}}$|

The random variable Y is different from X because it counts tails rather than heads. However, under $\widetilde{\mathbb{P}}$, the distribution of Y is the same as the distribution of X:

$$\begin{aligned} \widetilde{\mathbb{P}}\{Y = 0\} = \frac{1}{8}, \quad \widetilde{\mathbb{P}}\{Y = 1\} = \frac{3}{8},\\ \widetilde{\mathbb{P}}\{Y = 2\} = \frac{3}{8}, \quad \widetilde{\mathbb{P}}\{Y = 3\} = \frac{1}{8}. \end{aligned}$$

The point here is that the random variable is a function defined on Ω, whereas its distribution is a tabulation of probabilities that the random variable takes various values. A random variable is not a distribution.

Suppose, moreover, that we choose a probability measure $\mathbb{P}$ for Ω that corresponds to a $\frac{2}{3}$ probability of head on each toss and a $\frac{1}{3}$ probability of tail. Then

$$\begin{aligned} \mathbb{P}\{X = 0\} = \frac{1}{27}, \quad \mathbb{P}\{X = 1\} = \frac{6}{27},\\ \mathbb{P}\{X = 2\} = \frac{12}{27}, \quad \mathbb{P}\{X = 3\} = \frac{8}{27}. \end{aligned}$$

The random variable X has a different distribution under $\mathbb{P}$ than under $\widetilde{\mathbb{P}}$. It is the same random variable, counting the total number of heads, regardless of the probability measure used to determine its distribution. This is the situation we encounter later when we consider an asset price under both the actual and the risk-neutral probability measures.

Incidentally, although they have the same distribution under $\widetilde{\mathbb{P}}$, the random variables X and Y have different distributions under $\mathbb{P}$. Indeed,

$$\begin{aligned} \mathbb{P}\{Y = 0\} = \frac{8}{27}, \quad \mathbb{P}\{Y = 1\} = \frac{12}{27},\\ \mathbb{P}\{Y = 2\} = \frac{6}{27}, \quad \mathbb{P}\{Y = 3\} = \frac{1}{27}. \end{aligned}$$

□

Definition 2.2.4. *Let X be a random variable defined on a finite probability space $(\Omega, \mathbb{P})$. The* expectation *(or* expected value*) of X is defined to be*

$$\mathbb{E}X = \sum_{\omega \in \Omega} X(\omega)\mathbb{P}(\omega).$$

When we compute the expectation using the risk-neutral probability measure $\widetilde{\mathbb{P}}$, we use the notation

$$\widetilde{\mathbb{E}}X = \sum_{\omega \in \Omega} X(\omega)\widetilde{\mathbb{P}}(\omega).$$

The variance *of X is*

$$\mathrm{Var}(X) = \mathbb{E}\Big[(X - \mathbb{E}X^2\Big].$$

It is clear from its definition that expectation is linear: if X and Y are random variables and c_1 and c_2 are constants, then

$$\mathbb{E}(c_1X + c_2Y) = c_1\mathbb{E}X + c_2\mathbb{E}Y.$$

In particular, if $\ell(x) = ax+b$ is a linear function of a dummy variable x (a and b are constants), then $\mathbb{E}[\ell(X)] = \ell(\mathbb{E}X)$. When dealing with convex functions, we have the following inequality.

Theorem 2.2.5 (Jensen's inequality). *Let X be a random variable on a finite probability space, and let $\varphi(x)$ be a convex function of a dummy variable x. Then*

$$\mathbb{E}[\varphi(X)] \geq \varphi(\mathbb{E}X).$$

PROOF: We first argue that a convex function is the maximum of all linear functions that lie below it; i.e., for every $x \in \mathbb{R}$,

$$\varphi(x) = \max\{\ell(x); \ell \text{ is linear and } \ell(y) \leq \varphi(y) \text{ for all } y \in \mathbb{R}\}. \tag{2.2.1}$$

Since we are only considering linear functions that lie below φ, it is clear that

$$\varphi(x) \geq \max\{\ell(x); \ell \text{ is linear and } \ell(y) \leq \varphi(y) \text{ for all } y \in \mathbb{R}\}.$$

On the other hand, let x be an arbitrary point in $\mathbb{R}$. Because φ is convex, there is always a linear function ℓ that lies below φ and for which $\varphi(x) = \ell(x)$ for this particular x. This is called a *support line of φ at x* (see Figure 2.2.1). Therefore,

$$\varphi(x) \leq \max\{\ell(x); \ell \text{ is linear and } \ell(y) \leq \varphi(y) \text{ for all } y \in \mathbb{R}\}.$$

This establishes (2.2.1). Now let ℓ be a linear function lying below φ. We have

$$\mathbb{E}\left[\varphi(X)\right] \geq \mathbb{E}\left[\ell(X)\right] = \ell(\mathbb{E}X).$$

Since this inequality holds for every linear function ℓ lying below φ, we may take the maximum on the right-hand side over all such ℓ and obtain

$$\begin{aligned}\mathbb{E}\left[\varphi(X)\right] &\geq \max\{\ell(\mathbb{E}X); \ell \text{ is linear and } \ell(y) \leq \varphi(y) \text{ for all } y \in \mathbb{R}\}\\ &= \varphi(\mathbb{E}X).\end{aligned}$$

□

One consequence of Jensen's inequality is that

$$\mathbb{E}\left[X^2\right] \geq (\mathbb{E}X)^2.$$

We can also obtain this particular consequence of Jensen's inequality from the formula

$$0 \leq \mathbb{E}\Big[(X - \mathbb{E}X^2)\Big] = \mathbb{E}\Big[X^2 - 2X\mathbb{E}X + (\mathbb{E}X)^2\Big] = \mathbb{E}[X^2] - (\mathbb{E}X)^2.$$

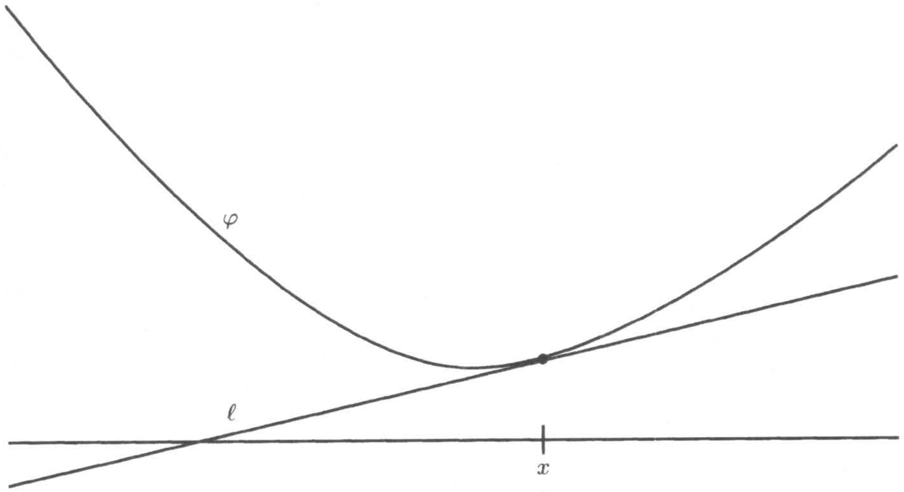

Fig. 2.2.1. Support line of φ at x.

2.3 Conditional Expectations

In the binomial pricing model of Chapter 1, we chose risk-neutral probabilities $\tilde{p}$ and $\tilde{q}$ by the formula (1.1.8), which we repeat here:

$$\tilde{p} = \frac{1+r-d}{u-d}, \quad \tilde{q} = \frac{u-1-r}{u-d}. \tag{2.3.1}$$

It is easily checked that these probabilities satisfy the equation

$$\frac{\tilde{p}u + \tilde{q}d}{1+r} = 1. \tag{2.3.2}$$

Consequently, at every time n and for every sequence of coin tosses $\omega_1 \dots \omega_n$, we have

$$S_n(\omega_1 \dots \omega_n) = \frac{1}{1+r}\Big[\tilde{p}S_{n+1}(\omega_1 \dots \omega_n H) + \tilde{q}S_{n+1}(\omega_1 \dots \omega_n T)\Big] \tag{2.3.3}$$

(i.e., the stock price at time n is the discounted weighted average of the two possible stock prices at time $n+1$, where $\tilde{p}$ and $\tilde{q}$ are the weights used in the averaging). To simplify notation, we define

$$\widetilde{\mathbb{E}}_n[S_{n+1}](\omega_1 \dots \omega_n) = \tilde{p}S_{n+1}(\omega_1 \dots \omega_n H) + \tilde{q}S_{n+1}(\omega_1 \dots \omega_n T) \tag{2.3.4}$$

so that we may rewrite (2.3.3) as

$$S_n = \frac{1}{1+r}\widetilde{\mathbb{E}}_n[S_{n+1}], \tag{2.3.5}$$

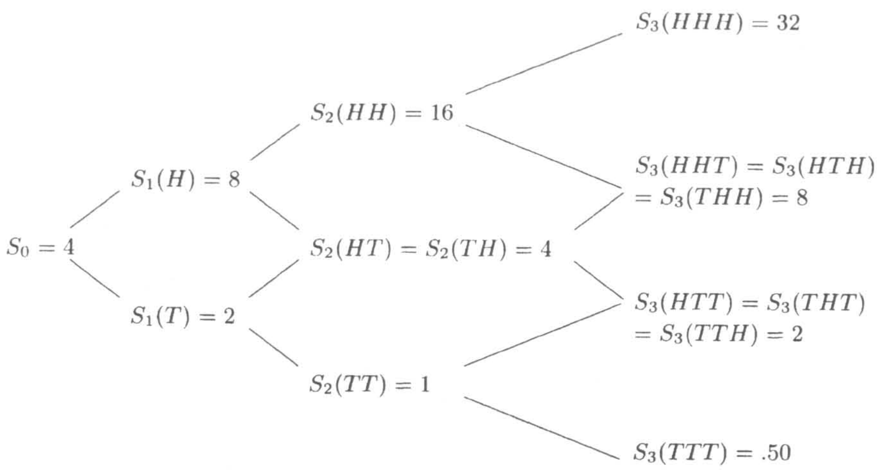

Fig. 2.3.1. A three-period model.

and we call $\widetilde{\mathbb{E}}_n[S_{n+1}]$ the *conditional expectation of* S_{n+1} *based on the information at time* n. The conditional expectation can be regarded as an estimate of the value of S_{n+1} based on knowledge of the first n coin tosses.

For example, in Figure 2.3.1 and using the risk-neutral probabilities $\tilde{p} = \tilde{q} = \frac{1}{2}$, we have $\widetilde{\mathbb{E}}_1[S_2](H) = 10$ and $\widetilde{\mathbb{E}}_1[S_2](T) = 2.50$. When we write simply $\widetilde{\mathbb{E}}_1[S_2]$ without specifying whether the first coin toss results in head or tail, we have a quantity whose value, not known at time zero, will be determined by the random experiment of coin tossing. According to Definition 2.2.1, such a quantity is a random variable.

More generally, whenever X is a random variable depending on the first N coin tosses, we can estimate X based on information available at an earlier time $n \le N$. The following definition generalizes (2.3.4).

Definition 2.3.1. *Let n satisfy $1 \le n \le N$, and let $\omega_1 \ldots \omega_n$ be given and, for the moment, fixed. There are 2^{N-n} possible continuations $\omega_{n+1} \ldots \omega_N$ of the sequence fixed $\omega_1 \ldots \omega_n$. Denote by $\#H(\omega_{n+1} \ldots \omega_N)$ the number of heads in the continuation $\omega_{n+1} \ldots \omega_N$ and by $\#T(\omega_{n+1} \ldots \omega_N)$ the number of tails. We define*

$$\begin{aligned}&\widetilde{\mathbb{E}}_n[X](\omega_1 \ldots \omega_n)\\ &= \sum_{\omega_{n+1}\ldots\omega_N} \tilde{p}^{\#H(\omega_{n+1}\ldots\omega_N)} \tilde{q}^{\#T(\omega_{n+1}\ldots\omega_N)} X(\omega_1 \ldots \omega_n \omega_{n+1} \ldots \omega_N) \qquad (2.3.6)\end{aligned}$$

and call $\widetilde{\mathbb{E}}_n[X]$ the conditional expectation of X based on the information at time n.

Based on what we know at time zero, the conditional expectation $\widetilde{\mathbb{E}}_n[X]$ is random in the sense that its value depends on the first n coin tosses, which we

do not know until time n. For example, in Figure 2.3.1 and using $\tilde{p} = \tilde{q} = \frac{1}{2}$, we obtain

$$\widetilde{\mathbb{E}}_1[S_3](H) = 12.50, \quad \widetilde{\mathbb{E}}_1[S_3](T) = 3.125,$$

so $\widetilde{\mathbb{E}}_1[S_3]$ is a random variable.

Definition 2.3.1 continued *The two extreme cases of conditioning are $\widetilde{\mathbb{E}}_0[X]$, the conditional expectation of X based on no information, which we define by*

$$\widetilde{\mathbb{E}}_0[X] = \widetilde{\mathbb{E}}X, \qquad (2.3.7)$$

and $\widetilde{\mathbb{E}}_N[X]$, the conditional expectation of X based on knowledge of all N coin tosses, which we define by

$$\widetilde{\mathbb{E}}_N[X] = X. \qquad (2.3.8)$$

The conditional expectations above have been computed using the risk-neutral probabilities $\tilde{p}$ and $\tilde{q}$. This is indicated by the $\tilde{}$ appearing in the notation $\widetilde{\mathbb{E}}_n$. Of course, conditional expectations can also be computed using the actual probabilities p and q, and these will be denoted by $\mathbb{E}_n$.

Regarded as random variables, conditional expectations have five fundamental properties, which we will use extensively. These are listed in the following theorem. We state them for conditional expectations computed under the actual probabilities, and the analogous results hold for conditional expectations computed under the risk-neutral probabilities.

Theorem 2.3.2 (Fundamental properties of conditional expectations). *Let N be a positive integer, and let X and Y be random variables depending on the first N coin tosses. Let $0 \le n \le N$ be given. The following properties hold.*

(i) ***Linearity of conditional expectations.*** *For all constants c_1 and c_2, we have*
$$\mathbb{E}_n[c_1X + c_2Y] = c_1\mathbb{E}_n[X] + c_2\mathbb{E}_n[Y].$$
(ii) ***Taking out what is known.*** *If X actually depends only on the first n coin tosses, then*
$$\mathbb{E}_n[XY] = X \cdot \mathbb{E}_n[Y].$$
(iii) ***Iterated conditioning.*** *If $0 \le n \le m \le N$, then*
$$\mathbb{E}_n\Big[\mathbb{E}_m[X]\Big] = \mathbb{E}_n[X].$$
In particular, $\mathbb{E}\Big[\mathbb{E}_m[X]\Big] = \mathbb{E}X$.

(iv) ***Independence.*** *If X depends only on tosses $n+1$ through N, then*
$$\mathbb{E}_n[X] = \mathbb{E}X.$$

(v) ***Conditional Jensen's inequality.*** *If $\varphi(x)$ is a convex function of the dummy variable x, then*

$$\mathbb{E}_n[\varphi(X)] \geq \varphi(\mathbb{E}_n[X]).$$

The proof of Theorem 2.3.2 is provided in the appendix. We illustrate the first four properties of the theorem with examples based on Figure 2.3.1 using the probabilities $p = \frac{2}{3}$, $q = \frac{1}{3}$. The fifth property, the conditional Jensen's inequality, follows from linearity of conditional expectations in the same way that Jensen's inequality for expectations follows from linearity of expectations (see the proof of Theorem 2.2.5).

Example 2.3.3 (Linearity of conditional expectations). With $p = \frac{2}{3}$ and $q = \frac{1}{3}$ in Figure 2.3.1, we compute

$$\mathbb{E}_1[S_2](H) = \frac{2}{3}\cdot 16 + \frac{1}{3}\cdot 4 = 12,$$
$$\mathbb{E}_1[S_3](H) = \frac{4}{9}\cdot 32 + \frac{2}{9}\cdot 8 + \frac{2}{9}\cdot 8 + \frac{1}{9}\cdot 2 = 18,$$

and consequently $\mathbb{E}_1[S_2](H) + \mathbb{E}_1[S_3](H) = 12 + 18 = 30$. But also

$$\mathbb{E}_1[S_2 + S_3](H) = \frac{4}{9}(16+32) + \frac{2}{9}(16+8) + \frac{2}{9}(4+8) + \frac{1}{9}(4+2) = 30.$$

A similar computation shows that

$$\mathbb{E}_1[S_2 + S_3](T) = 7.50 = \mathbb{E}_1[S_2](T) + \mathbb{E}_1[S_3](T).$$

In conclusion, regardless of the outcome of the first coin toss,

$$\mathbb{E}_1[S_2 + S_3] = \mathbb{E}_1[S_2] + E_1[S_3].$$

Example 2.3.4 (Taking out what is known). We first recall from Example 2.3.3 that

$$\mathbb{E}_1[S_2](H) = \frac{2}{3}\cdot 16 + \frac{1}{3}\cdot 4 = 12.$$

If we now want to estimate the product $S_1 S_2$ based on the information at time one, we can factor out the S_1, as seen by the following computation:

$$\mathbb{E}_1[S_1S_2](H) = \frac{2}{3}\cdot 128 + \frac{1}{3}\cdot 32 = 96 = 8\cdot 12 = S_1(H)\mathbb{E}_1[S_2](H).$$

A similar computation shows that

$$E_1[S_1S_2](T) = 6 = S_1(T)\mathbb{E}_1[S_2](T).$$

In conclusion, regardless of the outcome of the first toss,

$$\mathbb{E}_1[S_1S_2] = S_1\mathbb{E}_1[S_2].$$

Example 2.3.5 (Iterated conditioning). We first estimate S_3 based on the information at time two:

$$\mathbb{E}_2[S_3](HH) = \frac{2}{3}\cdot 32 + \frac{1}{3}\cdot 8 = 24,$$
$$\mathbb{E}_2[S_3](HT) = \frac{2}{3}\cdot 8 + \frac{1}{3}\cdot 2 = 6,$$
$$\mathbb{E}_2[S_3](TH) = \frac{2}{3}\cdot 8 + \frac{1}{3}\cdot 2 = 6,$$
$$\mathbb{E}_2[S_3](TT) = \frac{2}{3}\cdot 2 + \frac{1}{3}\cdot\frac{1}{2} = 1.50.$$

We now estimate the estimate, based on the information at time one:

$$\begin{aligned}\mathbb{E}_1\Big[\mathbb{E}_2[S_3]\Big](H) &= \frac{2}{3}\cdot\mathbb{E}_2[S_3](HH) + \frac{1}{3}\cdot\mathbb{E}_2[S_3](HT)\\ &= \frac{2}{3}\cdot 24 + \frac{1}{3}\cdot 6 = 18,\\ \mathbb{E}_1\Big[\mathbb{E}_2[S_3]\Big](T) &= \frac{2}{3}\cdot\mathbb{E}_2[S_3](TH) + \frac{1}{3}\mathbb{E}_2[S_3](TT)\\ &= \frac{2}{3}\cdot 6 + \frac{1}{3}\cdot 1.50 = 4.50.\end{aligned}$$

The estimate of the estimate is an average of averages, and it is not surprising that we can get the same result by a more comprehensive averaging. This more comprehensive averaging occurs when we estimate S_3 directly based on the information at time one:

$$E_1[S_3](H) = \frac{4}{9}\cdot 32 + \frac{2}{9}\cdot 8 + \frac{2}{9}\cdot 8 + \frac{1}{9}\cdot 2 = 18,$$
$$E_1[S_3](T) = \frac{4}{9}\cdot 8 + \frac{2}{9}\cdot 2 + \frac{2}{9}\cdot 2 + \frac{1}{9}\cdot\frac{1}{2} = 4.50.$$

In conclusion, regardless of the outcome of the first toss, we have

$$\mathbb{E}_1\Big[\mathbb{E}_2[S_3]\Big] = \mathbb{E}_1[S_3].$$

Example 2.3.6 (Independence). The quotient $\frac{S_2}{S_1}$ takes either the value 2 or $\frac{1}{2}$, depending on whether the second coin toss results in head or tail, respectively. In particular, $\frac{S_2}{S_1}$ does not depend on the first coin toss. We compute

$$\begin{aligned}\mathbb{E}_1\left[\frac{S_2}{S_1}\right](H) &= \frac{2}{3}\cdot\frac{S_2(HH)}{S_1(H)} + \frac{1}{3}\cdot\frac{S_2(HT)}{S_1(H)}\\ &= \frac{2}{3}\cdot 2 + \frac{1}{3}\cdot\frac{1}{2} = \frac{3}{2},\\ \mathbb{E}_1\left[\frac{S_2}{S_1}\right](T) &= \frac{2}{3}\cdot\frac{S_2(TH)}{S_1(T)} + \frac{1}{3}\cdot\frac{S_2(TT)}{S_1(T)}\\ &= \frac{2}{3}\cdot 2 + \frac{1}{3}\cdot\frac{1}{2} = \frac{3}{2}.\end{aligned}$$

We see that $\mathbb{E}_1\left[\frac{S_2}{S_1}\right]$ does not depend on the first coin toss (is not really random) and in fact is equal to

$$\mathbb{E}\frac{S_2}{S_1} = \frac{2}{3}\cdot 2 + \frac{1}{3}\cdot\frac{1}{2} = \frac{3}{2}.$$

2.4 Martingales

In the binomial pricing model of Chapter 1, we chose risk-neutral probabilities $\tilde{p}$ and $\tilde{q}$ so that at every time n and for every coin toss sequence $\omega_1 \dots \omega_n$ we have (2.3.3). In terms of the notation for conditional expectations introduced in Section 2.3, this fact can be written as (2.3.5). If we divide both sides of (2.3.5) by $(1+r)^n$, we get the equation

$$\frac{S_n}{(1+r)^n} = \widetilde{\mathbb{E}}_n\left[\frac{S_{n+1}}{(1+r)^{n+1}}\right]. \tag{2.4.1}$$

It does not matter in this model whether we write the term $\frac{1}{(1+r)^{n+1}}$ inside or outside the conditional expectation because it is constant (see Theorem 2.3.2(i)). In models with random interest rates, it would matter; we shall follow the practice of writing this term inside the conditional expectation since that is the way it would be written in models with random interest rates.

Equation (2.4.1) expresses the key fact that under the risk-neutral measure, for a stock that pays no dividend, the *best estimate based on the information at time n of the value of the discounted stock price at time n+1 is the discounted stock price at time n*. The risk-neutral probabilities are chosen to enforce this fact. Processes that satisfy this condition are called *martingales*. We give a formal definition of martingale under the actual probabilities p and q; the definition of martingale under the risk-neutral probabilities $\tilde{p}$ and $\tilde{q}$ is obtained by replacing $\mathbb{E}_n$ by $\widetilde{\mathbb{E}}_n$ in (2.4.2).

Definition 2.4.1. *Consider the binomial asset-pricing model. Let $M_0, M_1, \dots, M_N$ be a sequence of random variables, with each M_n depending only on the first n coin tosses (and M_0 constant). Such a sequence of random variables is called an* adapted stochastic process.

(i) If

$$M_n = \mathbb{E}_n[M_{n+1}],\ n = 0, 1, \dots, N-1, \tag{2.4.2}$$

we say this process is a martingale.

(ii) If

$$M_n \le \mathbb{E}_n[M_{n+1}],\ n = 0, 1, \dots, N-1,$$

we say the process is a submartingale *(even though it may have a tendency to increase);*

(iii) If

$$M_n \ge \mathbb{E}_n[M_{n+1}],\ n = 0, 1, \dots, N-1,$$

we say the process is a supermartingale *(even though it may have a tendency to decrease).*

Remark 2.4.2. The martingale property in (2.4.2) is a "one-step-ahead" condition. However, it implies a similar condition for any number of steps. Indeed, if $M_0, M_1, \dots, M_N$ is a martingale and $n \le N-2$, then the martingale property (2.4.2) implies

$$M_{n+1} = \mathbb{E}_{n+1}[M_{n+2}].$$

Taking conditional expectations on both sides based on the information at time n and using the iterated conditioning property (iii) of Theorem 2.3.2, we obtain

$$\mathbb{E}_n[M_{n+1}] = \mathbb{E}_n\Big[\mathbb{E}_{n+1}[M_{n+2}]\Big] = E_n[M_{n+2}].$$

Because of the martingale property (2.4.2), the left-hand side is M_n, and we thus have the "two-step-ahead" property

$$M_n = \mathbb{E}_n[M_{n+2}].$$

Iterating this argument, we can show that whenever $0 \le n \le m \le N$,

$$M_n = \mathbb{E}_n[M_m]. \tag{2.4.3}$$

One might call this the "multistep-ahead" version of the martingale property.

Remark 2.4.3. The expectation of a martingale is constant over time, i.e., if $M_0, M_1, \dots, M_N$ is a martingale, then

$$M_0 = \mathbb{E}M_n,\ n = 0, 1, \dots, N. \tag{2.4.4}$$

Indeed, if $M_0, M_1, \dots, M_N$ is a martingale, we may take expectations on both sides of (2.4.2), using Theorem 2.3.2(iii), and obtain $\mathbb{E}M_n = \mathbb{E}[M_{n+1}]$ for every n. It follows that

$$\mathbb{E}M_0 = \mathbb{E}M_1 = \mathbb{E}M_2 = \cdots = \mathbb{E}M_{N-1} = \mathbb{E}M_N.$$

But M_0 is not random, so $M_0 = \mathbb{E}M_0$, and (2.4.4) follows. □

In order to have a martingale, the equality in (2.4.2) must hold for all possible coin toss sequences. The stock price process in Figure 2.3.1 would be a martingale if the probability of an up move were $\hat{p} = \frac{1}{3}$ and the probability of a down move were $\hat{q} = \frac{2}{3}$ because, at every node in the tree in Figure 2.3.1, the stock price shown would then be the average of the two possible subsequent stock prices averaged with these weights. For example,

$$S_1(T) = 2 = \frac{1}{3}\cdot S_2(TH) + \frac{2}{3}\cdot S_2(TT).$$

A similar equation would hold at all other nodes in the tree, and therefore we would have a martingale under these probabilities.

A martingale has no tendency to rise or fall since the average of its next period values is always its value at the current time. Stock prices have a tendency to rise and, indeed, should rise on average faster than the money market in order to compensate investors for their inherent risk. In Figure 2.3.1 more realistic choices for p and q are $p = \frac{2}{3}$ and $q = \frac{1}{3}$. With these choices, we have

$$\mathbb{E}_n[S_{n+1}] = \frac{3}{2}S_n$$

at every node in the tree (i.e., on average, the next period stock price is 50% higher than the current stock price). This growth rate exceeds the 25% interest rate we have been using in this model, as it should. In particular, with $p = \frac{2}{3}$, $q = \frac{1}{3}$, and $r = \frac{1}{4}$, the discounted stock price has a tendency to rise. Note that when $r = \frac{1}{4}$, we have $\frac{1}{1+r} = \frac{4}{5}$, so the discounted stock price at time n is $\left(\frac{4}{5}\right)^n S_n$. We compute

$$\mathbb{E}_n\left[\left(\frac{4}{5}\right)^{n+1} S_{n+1}\right] = \left(\frac{4}{5}\right)^{n+1} \mathbb{E}_n[S_{n+1}] = \left(\frac{4}{5}\right)^{n} \cdot \frac{4}{5} \cdot \frac{3}{2} \cdot S_n \geq \left(\frac{4}{5}\right)^{n} S_n.$$

The discounted stock price is a submartingale under the actual probabilities $p = \frac{2}{3}$, $q = \frac{1}{3}$. This is typically the case in real markets.

The risk-neutral probabilities, on the other hand, are chosen to make the discounted stock price be a martingale. In Figure 2.3.1 with $\tilde{p} = \tilde{q} = \frac{1}{2}$, one can check that the martingale equation

$$\widetilde{\mathbb{E}}_n\left[\left(\frac{4}{5}\right)^{n+1} S_{n+1}\right] = \left(\frac{4}{5}\right)^{n} S_n \tag{2.4.5}$$

holds at every node. The following theorem shows that this example is representative.

Theorem 2.4.4. *Consider the general binomial model with* $0 < d < 1+r < u$. *Let the risk-neutral probabilities be given by*

$$\tilde{p} = \frac{1+r-d}{u-d}, \quad \tilde{q} = \frac{u-1-r}{u-d}.$$

Then, under the risk-neutral measure, the discounted stock price is a martingale, i.e., equation (2.4.1) holds at every time n and for every sequence of coin tosses.

We give two proofs of this theorem, an elementary one, which does not rely on Theorem 2.3.2, and a deeper one, which does rely on Theorem 2.3.2. The second proof will later be adapted to continuous-time models.

Note in Theorem 2.4.4 that the stock does not pay a dividend. For a dividend-paying stock, the situation is described in Exercise 2.10.

FIRST PROOF: Let n and $\omega_1 \ldots \omega_n$ be given. Then

$$\begin{aligned}
&\widetilde{\mathbb{E}}_n\left[\frac{S_{n+1}}{(1+r)^{n+1}}\right](\omega_1 \ldots \omega_n)\\
&\quad = \frac{1}{(1+r)^n} \cdot \frac{1}{1+r}\Big[\tilde{p}S_{n+1}(\omega_1 \ldots \omega_n H) + \tilde{q}S_{n-1}(\omega_1 \ldots \omega_n T)\Big]\\
&\quad = \frac{1}{(1+r)^n} \cdot \frac{1}{1+r}\Big[\tilde{p}uS_n(\omega_1 \ldots \omega_n) + \tilde{q}dS_n(\omega_1 \ldots \omega_n)\Big]\\
&\quad = \frac{S_n(\omega_1 \ldots \omega_n)}{(1+r)^n} \cdot \frac{\tilde{p}u + \tilde{q}d}{1+r}\\
&\quad = \frac{S_n(\omega_1 \ldots \omega_n)}{(1+r)^n}.
\end{aligned}$$

SECOND PROOF: Note that $\frac{S_{n+1}}{S_n}$ depends only on the $(n+1)$st coin toss. Using the indicated properties from Theorem 2.3.2, we compute

$$\begin{aligned}
\widetilde{\mathbb{E}}_n\left[\frac{S_{n+1}}{(1+r)^{n+1}}\right] &= \widetilde{\mathbb{E}}_n\left[\frac{S_n}{(1+r)^{n+1}} \cdot \frac{S_{n+1}}{S_n}\right]\\
&= \frac{S_n}{(1+r)^n}\widetilde{\mathbb{E}}_n\left[\frac{1}{1+r} \cdot \frac{S_{n+1}}{S_n}\right]\\
&\qquad \text{(Taking out what is known)}\\
&= \frac{S_n}{(1+r)^n} \cdot \frac{1}{1+r}\widetilde{\mathbb{E}}\frac{S_{n+1}}{S_n}\\
&\qquad \text{(Independence)}\\
&= \frac{S_n}{(1+r)^n} \frac{\tilde{p}u + \tilde{q}d}{1+r}\\
&= \frac{S_n}{(1+r)^n}.
\end{aligned}$$

□

In a binomial model with N coin tosses, we imagine an investor who at each time n takes a position of Δ_n shares of stock and holds this position until time $n+1$, when he takes a new position of Δ_{n+1} shares. The portfolio rebalancing at each step is financed by investing or borrowing, as necessary, from the money market. The "portfolio variable" Δ_n may depend on the first n coin tosses, and Δ_{n+1} may depend on the first $n+1$ coin tosses. In other words, the *portfolio process* $\Delta_0, \Delta_1, \ldots, \Delta_{N-1}$ is *adapted*, in the sense of Definition 2.4.1. If the investor begins with initial wealth X_0, and X_n denotes his wealth at each time n, then the evolution of his wealth is governed by the *wealth equation* (1.2.14) of Chapter 1, which we repeat here:

$$X_{n+1} = \Delta_n S_{n+1} + (1+r)(X_n - \Delta_n S_n), \; n = 0, 1, \ldots, N-1. \tag{2.4.6}$$

Note that each X_n depends only on the first n coin tosses (i.e., the *wealth process* is adapted).

We may inquire about the average rate of growth of the investor's wealth. If we mean the average under the actual probabilities, the answer depends on the portfolio process he uses. In particular, since a stock generally has a higher average rate of growth than the money market, the investor can achieve a rate of growth higher than the interest rate by taking long positions in the stock. Indeed, by borrowing from the money market, the investor can achieve an arbitrarily high *average* rate of growth. Of course, such leveraged positions are also extremely risky.

On the other hand, if we want to know the average rate of growth of the investor's wealth under the risk-neutral probabilities, the portfolio the investor uses is irrelevant. Under the risk-neutral probabilities, the average rate of growth of the stock is equal to the interest rate. No matter how the investor divides his wealth between the stock and the money market account, he will achieve an average rate of growth equal to the interest rate. Although some portfolio processes are riskier then others under the risk-neutral measure, they all have the same average rate of growth. We state this result as a theorem, whose proof is given in a way that we can later generalize to continuous time.

Theorem 2.4.5. *Consider the binomial model with N periods. Let $\Delta_0, \Delta_1, \dots, \Delta_{N-1}$ be an adapted portfolio process, let X_0 be a real number, and let the wealth process $X_1, \dots, X_N$ be generated recursively by (2.4.6). Then the discounted wealth process $\frac{X_n}{(1+r)^n}$, $n = 0, 1, \dots, N$, is a martingale under the risk-neutral measure; i.e.,*

$$\frac{X_n}{(1+r)^n} = \widetilde{\mathbb{E}}_n\left[\frac{X_{n+1}}{(1+r)^{n+1}}\right], \quad n = 0, 1, \dots, N-1. \tag{2.4.7}$$

PROOF: We compute

$$\begin{aligned}
\widetilde{\mathbb{E}}_n\left[\frac{X_{n+1}}{(1+r)^{n+1}}\right] &= \widetilde{\mathbb{E}}_n\left[\frac{\Delta_n S_{n+1}}{(1+r)^{n+1}} + \frac{X_n - \Delta_n S_n}{(1+r)^n}\right]\\
&= \widetilde{\mathbb{E}}_n\left[\frac{\Delta_n S_{n+1}}{(1+r)^{n+1}}\right] + \widetilde{\mathbb{E}}_n\left[\frac{X_n - \Delta_n S_n}{(1+r)^n}\right]\\
&\quad \text{(Linearity)}\\
&= \Delta_n \widetilde{\mathbb{E}}_n\left[\frac{S_{n+1}}{(1+r)^{n+1}}\right] + \frac{X_n - \Delta_n S_n}{(1+r)^n}\\
&\quad \text{(Taking out what is known)}\\
&= \Delta_n \frac{S_n}{(1+r)^n} + \frac{X_n - \Delta_n S_n}{(1+r)^n}\\
&\quad \text{(Theorem 2.4.4)}\\
&= \frac{X_n}{(1+r)^n}.
\end{aligned}$$

□

Corollary 2.4.6. *Under the conditions of Theorem 2.4.5, we have*

$$\widetilde{\mathbb{E}}\frac{X_n}{(1+r)^n} = X_0, \ n = 0, 1, \dots, N. \tag{2.4.8}$$

PROOF: The corollary follows from the fact that the expected value of a martingale cannot change with time and so must always be equal to the time-zero value of the martingale (see Remark 2.4.3). Applying this fact to the $\widetilde{\mathbb{P}}$-martingale $\frac{X_n}{(1+r)^n}$, $n = 0, 1, \dots, N$, we obtain (2.4.8). □

Theorem 2.4.5 and its corollary have two important consequences. The first is that there can be no arbitrage in the binomial model. If there were an arbitrage, we could begin with $X_0 = 0$ and find a portfolio process whose corresponding wealth process $X_1, X_2, \dots, X_N$ satisfied $X_N(\omega) \geq 0$ for all coin toss sequences ω and $X_N(\overline{\omega}) > 0$ for at least one coin toss sequence $\overline{\omega}$. But then we would have $\widetilde{\mathbb{E}}X_0 = 0$ and $\widetilde{\mathbb{E}}\frac{X_N}{(1+r)^N} > 0$, which violates Corollary 2.4.6.

In general, if we can find a risk-neutral measure in a model (i.e., a measure that agrees with the actual probability measure about which price paths have zero probability, and under which the discounted prices of all primary assets are martingales), then there is no arbitrage in the model. This is sometimes called the *First Fundamental Theorem of Asset Pricing*. The essence of its proof is contained in the preceding paragraph: under a risk-neutral measure, the discounted wealth process has constant expectation, so it cannot begin at zero and later be strictly positive with positive probability unless it also has a positive probability of being strictly negative. The First Fundamental Theorem of Asset Pricing will prove useful for ruling out arbitrage in term-structure models later on and thereby lead to the Heath-Jarrow-Morton no-arbitrage condition on forward rates.

The other consequence of Theorem 2.4.5 is the following version of the *risk-neutral pricing formula*. Let V_N be a random variable (derivative security paying off at time N) depending on the first N coin tosses. We know from Theorem 1.2.2 of Chapter 1 that there is an initial wealth X_0 and a replicating portfolio process $\Delta_0, \dots, \Delta_{N-1}$ that generates a wealth process $X_1, \dots, X_N$ satisfying $X_N = V_N$, no matter how the coin tossing turns out. Because $\frac{X_n}{(1+r)^n}$, $n = 0, 1, \dots, N$, is a martingale, the "multistep ahead" property of Remark 2.4.2 implies

$$\frac{X_n}{(1+r)^n} = \widetilde{\mathbb{E}}_n\left[\frac{X_N}{(1+r)^N}\right] = \widetilde{\mathbb{E}}_n\left[\frac{V_N}{(1+r)^N}\right]. \tag{2.4.9}$$

According to Definition 1.2.3 of Chapter 1, we define the price of the derivative security at time n to be X_n and denote this price by the symbol V_n. Thus, (2.4.9) may be rewritten as

$$\frac{V_n}{(1+r)^n} = \widetilde{\mathbb{E}}_n\left[\frac{V_N}{(1+r)^N}\right] \tag{2.4.10}$$

or, equivalently,

$$V_n = \widetilde{\mathbb{E}}_n \left[\frac{V_N}{(1+r)^{N-n}} \right]. \tag{2.4.11}$$

We summarize with a theorem.

Theorem 2.4.7 (Risk-neutral pricing formula). *Consider an N-period binomial asset-pricing model with $0 < d < 1 + r < u$ and with risk-neutral probability measure $\widetilde{\mathbb{P}}$. Let V_N be a random variable (a derivative security paying off at time N) depending on the coin tosses. Then, for n between 0 and N, the price of the derivative security at time n is given by the* risk-neutral pricing formula *(2.4.11). Furthermore, the discounted price of the derivative security is a martingale under $\widetilde{\mathbb{P}}$; i.e.,*

$$\frac{V_n}{(1+r)^n} = \widetilde{\mathbb{E}}_n \left[\frac{V_{n+1}}{(1+r)^{n+1}} \right], \quad n = 0, 1, \ldots, N-1. \tag{2.4.12}$$

The random variables V_n defined by (2.4.11) are the same as the random variable V_n defined in Theorem 1.2.2.

The remaining steps in the proof of Theorem 2.4.7 are outlined in Exercise 2.8. We note that we chose the risk-neutral measure in order to make the discounted stock price a martingale. According to Theorem 2.4.7, a consequence of this is that discounted derivative security prices under the risk-neutral measure are also martingales.

So far, we have discussed only derivative securities that pay off on a single date. Many securities, such as coupon-paying bonds and interest rate swaps, make a series of payments. For such a security, we have the following pricing and hedging formulas.

Theorem 2.4.8 (Cash flow valuation). *Consider an N-period binomial asset pricing-model with $0 < d < 1 + r < u$, and with risk-neutral probability measure $\widetilde{\mathbb{P}}$. Let $C_0, C_1, \ldots, C_N$ be a sequence of random variables such that each C_n depends only on $\omega_1 \ldots \omega_n$. The price at time n of the derivative security that makes payments $C_n, \ldots, C_N$ at times $n, \ldots, N$, respectively, is*

$$V_n = \widetilde{\mathbb{E}}_n \left[\sum_{k=n}^{N} \frac{C_k}{(1+r)^{k-n}} \right], \quad n = 0, 1, \ldots, N. \tag{2.4.13}$$

The price process V_n, $n = 0, 1, \ldots, N$, satisfies

$$C_n(\omega_1 \ldots \omega_n) = V_n(\omega_1 \ldots \omega_n) - \frac{1}{1+r} \Big[\tilde{p} V_{n+1}(\omega_1 \ldots \omega_n H) + \tilde{q} V_{n+1}(\omega_1 \ldots \omega_n T) \Big]. \tag{2.4.14}$$

We define

$$\Delta_n(\omega_1 \ldots \omega_n) = \frac{V_{n+1}(\omega_1 \ldots \omega_n H) - V_{n+1}(\omega_1 \ldots \omega_n T)}{S_{n+1}(\omega_1 \ldots \omega_n H) - S_{n+1}(\omega_1 \ldots \omega_n T)}, \tag{2.4.15}$$

where n ranges between 0 and $N-1$. If we set $X_0 = V_0$ and define recursively forward in time the portfolio values $X_1, X_2, \ldots, X_N$ by

$$X_{n+1} = \Delta_n S_{n+1} + (1+r)(X_n - C_n - \Delta_n S_n), \tag{2.4.16}$$

then we have

$$X_n(\omega_1 \ldots \omega_n) = V_n(\omega_1 \ldots \omega_n) \tag{2.4.17}$$

for all n and all $\omega_1 \ldots \omega_n$.

In Theorem 2.4.8, V_n is the so-called *net present value* at time n of the sequence of payments $C_n, \ldots, C_N$. It is just the sum of the value $\widetilde{\mathbb{E}}_n \left[\frac{C_k}{(1+r)^{(n-k)}} \right]$ of each of the payments C_k to be made at times $k = n, k = n+1, \ldots, k = N$. Note that the payment at time n is included. This payment C_n depends on only the first n tosses and so can be taken outside the conditional expectation $\widetilde{\mathbb{E}}_n$, i.e.,

$$V_n = C_n + \widetilde{\mathbb{E}}_n \left[\sum_{k=n+1}^{N} \frac{C_k}{(1+r)^{k-n}} \right], n = 0, 1, \ldots, N-1. \tag{2.4.18}$$

In the case of $n = N$, (2.4.13) reduces to

$$V_N = C_N. \tag{2.4.19}$$

Consider an agent who is short the cash flows represented by $C_0, \ldots, C_n$ (i.e., an agent who must make the payment C_n at each time n). (We allow these payments to be negative as well as positive. If a payment is negative, the agent who is short actually receives cash.) Suppose the agent in the short position invests in the stock and money market account, so that, at time n, before making the payment C_n, the value of his portfolio is X_n. He then makes the payment C_n. Suppose he then takes a position Δ_n in stock. This will cause the value of his portfolio at time $n+1$ before making the payment C_{n+1} to be X_{n+1}, given by (2.4.16). If this agent begins with $X_0 = V_0$ and chooses his stock positions Δ_n by (2.4.15), then (2.4.17) holds and, in particular, $X_N = V_N = C_N$ (see (2.4.17), and (2.4.19)). Then, at time N he makes the final payment C_N and is left with 0. He has perfectly hedged the short position in the cash flows. This is the justification for calling V_n the value at time n of the future cash flows, including the payment C_n to be made at time n.

PROOF OF THEOREM 2.4.8: To prove (2.4.17), we proceed by induction on n. The induction hypothesis is that $X_n(\omega_1 \ldots \omega_n) = V_n(\omega_1 \ldots \omega_n)$ for some $n \in \{0, 1, \ldots, N-1\}$ and all $\omega_1 \ldots \omega_n$. We need to show that

$$X_{n+1}(\omega_1 \ldots \omega_n H) = V_{n+1}(\omega_1 \ldots \omega_n H), \tag{2.4.20}$$

$$X_{n+1}(\omega_1 \ldots \omega_n T) = V_{n+1}(\omega_1 \ldots \omega_n T). \tag{2.4.21}$$

We prove (2.4.20); the proof of (2.4.21) is analogous.

From (2.4.18) and iterated conditioning (Theorem 2.3.2(iii)), we have

$$\begin{aligned} V_n &= C_n + \widetilde{\mathbb{E}}_n \left[\frac{1}{1+r} \widetilde{\mathbb{E}}_{n+1} \left[\sum_{k=n+1}^{N} \frac{C_k}{(1+r)^{k-(n+1)}} \right] \right] \\ &= C_n + \widetilde{\mathbb{E}}_n \left[\frac{1}{1+r} V_{n+1} \right], \end{aligned}$$

where we have used (2.4.13) with n replaced by $n+1$ in the last step. In other words, for all $\omega_1 \dots \omega_n$, we have

$$\begin{aligned} &V_n(\omega_1 \dots \omega_n) - C_n(\omega_1 \dots \omega_n) \\ &= \frac{1}{1+r} \big[\tilde{p} V_{n+1}(\omega_1 \dots \omega_n H) + \tilde{q} V_{n+1}(\omega_1 \dots \omega_n T) \big]. \end{aligned}$$

Since $\omega_1 \dots \omega_n$ will be fixed for the rest of the proof, we will suppress these symbols. For example, the last equation will be written simply as

$$V_n - C_n = \frac{1}{1+r} \big[\tilde{p} V_{n+1}(H) + \tilde{q} V_{n+1}(T) \big].$$

We compute

$$\begin{aligned} X_{n+1}(H) &= \Delta_n S_{n+1}(H) + (1+r)(X_n - C_n - \Delta_n S_n) \\ &= \frac{V_{n+1}(H) - V_{n+1}(T)}{S_{n+1}(H) - S_{n+1}(T)} (S_{n+1}(H) - (1+r)S_n) \\ &\quad + (1+r)(V_n - C_n) \\ &= \frac{V_{n+1}(H) - V_{n+1}(T)}{(u-d)S_n} (uS_n - (1+r)S_n) \\ &\quad + \tilde{p} V_{n+1}(H) + \tilde{q} V_{n+1}(T) \\ &= \big(V_{n+1}(H) - V_{n+1}(T) \big) \frac{u-1-r}{u-d} + \tilde{p} V_{n+1}(H) + \tilde{q} V_{n+1}(T) \\ &= \big(V_{n+1}(H) - V_{n+1}(T) \big) \tilde{q} + \tilde{p} V_{n+1}(H) + \tilde{q} V_{n+1}(T) \\ &= (\tilde{p} + \tilde{q}) V_{n+1}(H) \;=\; V_{n+1}(H). \end{aligned}$$

This is (2.4.20). □

2.5 Markov Processes

In Section 1.3, we saw that the computational requirements of the derivative security pricing algorithm of Theorem 1.2.2 can often be substantially reduced by thinking carefully about what information needs to be remembered as we go from period to period. In Example 1.3.1 of Section 1.3, the stock price was relevant, but the path it followed to get to its current price was not. In Example 1.3.2 of Section 1.3, the stock price and the maximum value it had achieved up to the current time were relevant. In this section, we formalize the procedure for determining what is relevant and what is not.

Definition 2.5.1. *Consider the binomial asset-pricing model. Let $X_0, X_1, \dots, X_N$ be an adapted process. If, for every n between 0 and $N-1$ and for every function $f(x)$, there is another function $g(x)$ (depending on n and f) such that*

$$\mathbb{E}_n[f(X_{n+1})] = g(X_n), \tag{2.5.1}$$

we say that $X_0, X_1, \dots, X_N$ is a Markov process.

By definition, $\mathbb{E}_n[f(M_{n+1})]$ is random; it depends on the first n coin tosses. The Markov property says that this dependence on the coin tosses occurs through X_n (i.e., the information about the coin tosses one needs in order to evaluate $\mathbb{E}_n[f(X_{n+1})]$ is summarized by X_n). We are not so concerned with determining a formula for the function g right now as we are with asserting its existence because its mere existence tells us that if the payoff of a derivative security is random only through its dependence on X_N, then there is a version of the derivative security pricing algorithm in which we do not need to store path information (see Theorem 2.5.8). In examples in this section, we shall develop a method for finding the function g.

Example 2.5.2 (Stock price). In the binomial model, the stock price at time $n+1$ is given in terms of the stock price at time n by the formula

$$S_{n+1}(\omega_1 \dots \omega_n \omega_{n+1}) = \begin{cases} uS_n(\omega_1 \dots \omega_n), & \text{if } \omega_{n+1} = H, \\ dS_n(\omega_1 \dots \omega_n), & \text{if } \omega_{n+1} = T. \end{cases}$$

Therefore,

$$\mathbb{E}_n[f(S_{n+1})](\omega_1 \dots \omega_n) = pf(uS_n(\omega_1 \dots \omega_n)) + qf(dS_n(\omega_1 \dots \omega_n)),$$

and the right-hand side depends on $\omega_1 \dots \omega_n$ only through the value of $S_n(\omega_1 \dots \omega_n)$. Omitting the coin tosses $\omega_1 \dots \omega_n$, we can rewrite this equation as

$$\mathbb{E}_n[f(S_{n+1})] = g(S_n),$$

where the function $g(x)$ of the dummy variable x is defined by $g(x) = pf(ux) + qf(dx)$. This shows that the stock price process is Markov.

Indeed, the stock price process is Markov under either the actual or the risk-neutral probability measure. To determine the price V_n at time n of a derivative security whose payoff at time N is a function v_N of the stock price S_N (i.e., $V_N = v_N(S_N)$), we use the risk-neutral pricing formula (2.4.12), which reduces to

$$V_n = \frac{1}{1+r} \widetilde{\mathbb{E}}_n[V_{n+1}], \quad n = 0, 1, \dots, N-1.$$

But $V_N = v_N(S_N)$ and the stock price process is Markov, so

$$V_{N-1} = \frac{1}{1+r} \widetilde{\mathbb{E}}_{N-1}[v_N(S_N)] = v_{N-1}(S_{N-1})$$

for some function v_{N-1}. Similarly,

$$V_{N-2} = \frac{1}{1+r}\widetilde{\mathbb{E}}_{N-2}[v_{N-1}(S_{N-1})] = v_{N-2}(S_{N-2})$$

for some function v_{N-2}. In general, $V_n = v_n(S_n)$ for some function v_n. Moreover, we can compute these functions recursively by the algorithm

$$v_n(s) = \frac{1}{1+r}\Big[\tilde{p}v_{n+1}(us) + \tilde{q}v_{n+1}(ds)\Big],\ n = N-1, N-2, \dots, 0. \quad (2.5.2)$$

This algorithm works in the binomial model for any derivative security whose payoff at time N is a function only of the stock price at time N. In particular, we have the same algorithm for puts and calls. The only difference is in the formula for $v_N(s)$. For the call, we have $v_N(s) = (s-K)^+$; for the put, we have $v_N(s) = (K-s)^+$. □

The martingale property is the special case of (2.5.1) with $f(x) = x$ and $g(x) = x$. In order for a process to be Markov, it is necessary that for *every* function f there must be a corresponding function g such that (2.5.1) holds. Not every martingale is Markov. On the other hand, even when considering the function $f(x) = x$, the Markov property requires only that $\mathbb{E}_n[M_{n+1}] = g(M_n)$ for some function g; it does not require that the function g be given by $g(x) = x$. Not every Markov process is a martingale. Indeed, Example 2.5.2 shows that the stock price is Markov under both the actual and the risk-neutral probability measures. It is typically not a martingale under either of these measures. However, if $pu + qd = 1$, then the stock price is both a martingale and a Markov process under the actual probability measure.

The following lemma often provides the key step in the verification that a process is Markov.

Lemma 2.5.3 (Independence). *In the N-period binomial asset pricing model, let n be an integer between 0 and N. Suppose the random variables $X^1, \dots, X^K$ depend only on coin tosses 1 through n and the random variables $Y^1, \dots, Y^L$ depend only on coin tosses $n+1$ through N. (The superscripts $1, \dots, K$ on X and $1, \dots, L$ on Y are superscripts, not exponents.) Let $f(x^1, \dots, x^K, y^1, \dots, y^L)$ be a function of dummy variables $x^1, \dots, x^K$ and $y^1, \dots, y^L$, and define*

$$g(x^1, \dots, x^K) = \mathbb{E}f(x^1, \dots, x^K, Y^1, \dots, Y^L). \quad (2.5.3)$$

Then

$$\mathbb{E}_n[f(X^1, \dots, X^K, Y^1, \dots, Y^L)] = g(X^1, \dots, X^K). \quad (2.5.4)$$

For the following discussion and proof of the lemma, we assume that $K = L = 1$. Then (2.5.3) takes the form

$$g(x) = \mathbb{E}f(x, Y) \quad (2.5.3)'$$

and (2.5.4) takes the form

$$\mathbb{E}_n[f(X, Y)] = g(X), \quad (2.5.4)'$$

where the random variable X is assumed to depend only on the first n coin tosses, and the random variable Y depends only on coin tosses $n+1$ through N.

This lemma generalizes the property "taking out what is known" of Theorem 2.3.2(ii). Since X is "known" at time n, we want to "take it out" of the computation of the conditional expectation $\mathbb{E}_n[f(X, Y)]$. However, because X is inside the argument of the function f, we cannot simply factor it out as we did in Theorem 2.3.2(ii). Therefore, we hold it constant by replacing the random variable X by an arbitrary but fixed dummy variable x. We then compute the conditional expectation of the random variable $f(x, Y)$, whose randomness is due only to the dependence of Y on tosses $n+1$ through N. Because of Theorem 2.3.2(iv), this conditional expectation is the same as the unconditional expectation in (2.5.3)′. Finally, we recall that $E_n[f(X, Y)]$ must depend on the value of the random variable X, so we replace the dummy variable x by the random variable X after g is computed.

In the context of Example 2.5.2, we can take $X = S_n$, which depends only on the first n coin tosses, and take $Y = \frac{S_{n+1}}{S_n}$, which depends only on the $(n+1)$st coin toss, taking the value u if the $(n+1)$st toss results in a head and taking the value d if it results in a tail. We are asked to compute

$$\mathbb{E}_n[f(S_{n+1})] = \mathbb{E}_n[f(XY)].$$

We replace X by a dummy variable x and compute

$$g(x) = \mathbb{E}f(xY) = pf(ux) + qf(dx).$$

Then $\mathbb{E}_n[f(S_{n+1}] = g(S_n)$.

PROOF OF LEMMA 2.5.3: Let $\omega_1 \dots \omega_n$ be fixed but arbitrary. By the definition (2.3.6) of conditional expectation,

$$\begin{aligned}&\mathbb{E}_n[f(X, Y)](\omega_1 \dots \omega_n)\\ &= \sum_{\omega_{n+1}\dots\omega_N} f(X(\omega_1 \dots \omega_n), Y(\omega_{n+1} \dots \omega_N))p^{\#H(\omega_{n+1}\dots\omega_N)}q^{\#T(\omega_{n+1}\dots\omega_N)},\end{aligned}$$

whereas

$$\begin{aligned}g(x) &= \mathbb{E}f(x, Y)\\ &= \sum_{\omega_{n+1}\dots\omega_N} f(x, Y(\omega_{n+1} \dots \omega_N))p^{\#H(\omega_{n+1}\dots\omega_N)}q^{\#T(\omega_{n+1}\dots\omega_N)}.\end{aligned}$$

It is apparent that

$$\mathbb{E}_n[f(X, Y)](\omega_1 \dots \omega_N) = g(X(\omega_1 \dots \omega_N)).$$ □

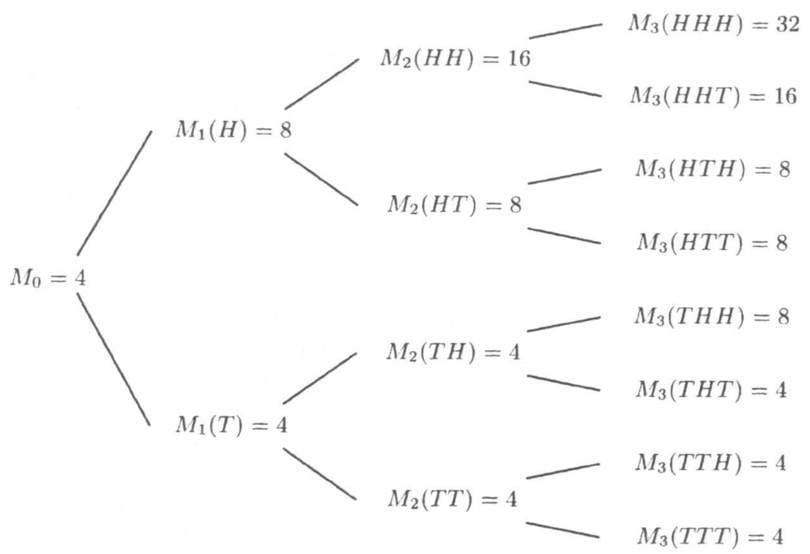

Fig. 2.5.1. The maximum stock price to date.

Example 2.5.4 (Non-Markov process). In the binomial model of Figure 2.3.1, consider the maximum-to-date process $M_n = \max_{0 \le k \le n} S_k$, shown in Figure 2.5.1. With $p = \frac{2}{3}$ and $q = \frac{1}{3}$, we have

$$\mathbb{E}_2[M_3](TH) = \frac{2}{3}M_3(THH) + \frac{1}{3}M_3(THT) = \frac{16}{3} + \frac{4}{3} = 6\frac{2}{3},$$

but

$$\mathbb{E}_2[M_3](TT) = \frac{2}{3}M_3(TTH) + \frac{1}{3}M_3(TTT) = \frac{8}{3} + \frac{4}{3} = 4.$$

Since $M_2(TH) = M_2(TT) = 4$, there cannot be a function g such that $\mathbb{E}_3[M_3](TH) = g(M_2(TH))$ and $\mathbb{E}_3[M_3](TT) = g(M_2(TT))$. The right-hand sides would be the same, but the left-hand sides would not. The maximum-to-date process is not Markov because recording only that the value of the maximum-to-date at time two is 4, without recording the value of the stock price at time two, neglects information relevant to the evolution of the maximum-to-date process after time two. □

When we encounter a non-Markov process, we can sometimes recover the Markov property by adding one or more so-called *state variables*. The term "state variable" is used because if we can succeed in recovering the Markov property by adding these variables, we will have determined a way to describe the "state" of the market in terms of these variables. This approach to recovering the Markov property requires that we generalize Definition 2.5.1 to multidimensional processes.

Definition 2.5.5. *Consider the binomial asset-pricing model. Let* $\{(X_n^1, \dots, X_n^K); n = 0, 1, \dots, N\}$ *be a K-dimensional adapted process; i.e., K one-dimensional adapted processes. If, for every n between 0 and $N-1$ and for every function $f(x^1, \dots, x^K)$, there is another function $g(x^1, \dots, x^K)$ (depending on n and f) such that*

$$\mathbb{E}_n[f(X_{n+1}^1, \dots, X_{n+1}^K)] = g(X_n^1, \dots, X_n^K), \tag{2.5.5}$$

we say that $\{(X_n^1, \dots, X_n^K); n = 0, 1, \dots, N\}$ *is a K-dimensional* Markov process.

Example 2.5.6. In an N-period binomial model, consider the two-dimensional adapted process $\{(S_n, M_n); n = 0, 1, \dots, N\}$, where S_n is the stock price at time n and $M_n = \max_{0 \le k \le n} S_k$ is the stock price maximum-to-date. We show that this two-dimensional process is Markov. To do that, we define $Y = \frac{S_{n+1}}{S_n}$, which depends only on the $(n+1)$st coin toss. Then

$$S_{n+1} = S_n Y$$

and

$$M_{n+1} = M_n \vee S_{n+1} = M_n \vee (S_n Y),$$

where $x \vee y = \max\{x, y\}$. We wish to compute

$$\mathbb{E}_n[f(S_{n+1}, M_{n+1})] = \mathbb{E}_n[f(S_n Y, M_n \vee (S_n Y))].$$

According to Lemma 2.5.3, we replace S_n by a dummy variable s, replace M_n by a dummy variable m, and compute

$$g(s, m) = \mathbb{E} f(sY, m \vee (sY)) = pf(us, m \vee (us)) + qf(ds, m \vee (ds)).$$

Then

$$\mathbb{E}_n[f(S_{n+1}, M_{n+1})] = g(S_n, M_n).$$

Since we have obtained a formula for $\mathbb{E}_n[f(S_{n+1}, M_{n+1})]$ in which the only randomness enters through the random variables S_n and M_n, we conclude that the two-dimensional process is Markov. In this example, we have used the actual probability measure, but the same argument shows that $\{(S_n, M_n); n = 0, 1, \dots, N\}$ is Markov under the risk-neutral probability measure $\widetilde{\mathbb{P}}$. □

Remark 2.5.7. The Markov property, in both the one-dimensional form of Definition 2.5.1 and the multidimensional form of Definition 2.5.5, is a "one-step-ahead" property, determining a formula for the conditional expectation of X_{n+1} in terms of X_n. However, it implies a similar condition for any number of steps. Indeed, if $X_0, X_1, \dots, X_N$ is a Markov process and $n \le N-2$, then the "one-step-ahead" Markov property implies that for every function h there is a function f such that

$$\mathbb{E}_{n+1}[h(X_{n+2})] = f(X_{n+1}).$$

Taking conditional expectations on both sides based on the information at time n and using the iterated conditioning property (iii) of Theorem 2.3.3, we obtain

$$\mathbb{E}_n[h(X_{n+2})] = \mathbb{E}_n\Big[\mathbb{E}_{n+1}[h(X_{n+2})]\Big] = \mathbb{E}_n[f(X_{n+1})].$$

Because of the "one-step-ahead" Markov property, the right-hand side is $g(X_n)$ for some function g, and we have obtained the "two-step-ahead" Markov property

$$\mathbb{E}_n[h(X_{n+2})] = g(X_n).$$

Iterating this argument, we can show that whenever $0 \le n \le m \le N$ and h is any function, then there is another function g such that the "multi-step-ahead" Markov property

$$\mathbb{E}_n[h(X_m)] = g(X_n) \tag{2.5.6}$$

holds. Similarly, if $\{(X_n^1, \dots, X_n^K); n = 1, 2, \dots, N\}$ is a K-dimensional Markov process, then whenever $0 \le n \le m \le N$ and $h(x^1, \dots, x^K)$ is any function, there is another function $g(x^1, \dots, x^K)$ such that

$$\mathbb{E}_n[h(X_m^1, \dots, X_m^K)] = g(X_n^1, \dots, X_n^K). \tag{2.5.7}$$

□

In the binomial pricing model, suppose we have a Markov process $X_0, X_1, \dots, X_N$ under the risk-neutral probability measure $\widetilde{\mathbb{P}}$, and we have a derivative security whose payoff V_N at time N is a function v_N of X_N, i.e., $V_N = v_N(X_N)$. The difference between V_N and v_N is that the argument of the former is $\omega_1 \dots \omega_N$, a sequence of coin tosses, whereas the argument of the latter is a real number, which we will sometimes denote by the dummy variable x. In particular, there is nothing random about $v_N(x)$. However, if in place of the dummy variable x we substitute the random variable X_N (actually $X_N(\omega_1 \dots \omega_N)$), then we have a random variable. Indeed, we have

$$V_N(\omega_1 \dots \omega_N) = v_N(X_N(\omega_1 \dots \omega_N)) \text{ for all } \omega_1 \dots \omega_N.$$

The risk-neutral pricing formula (2.4.11) says that the price of this derivative security at earlier times n is

$$V_n(\omega_1 \dots \omega_n) = \widetilde{\mathbb{E}}_n\left[\frac{V_N}{(1+r)^{N-n}}\right](\omega_1 \dots \omega_n) \text{ for all } \omega_1 \dots \omega_n.$$

On the other hand, the "multi-step-ahead" Markov property implies that there is a function v_n such that

$$\widetilde{\mathbb{E}}_n\left[\frac{V_N}{(1+r)^{N-n}}\right](\omega_1 \dots \omega_n) = v_n(X_n(\omega_1 \dots \omega_n)) \text{ for all } \omega_1 \dots \omega_n.$$

Therefore, the price of the derivative security at time n is a function of X_n, i.e.,

$$V_n = v_n(X_n).$$

Instead of computing the random variables V_n, we can compute the functions v_n, and this is generally much more manageable computationally. In particular, when the Markov process $X_0, X_1, \dots, X_N$ is the stock price itself, we get the algorithm (2.5.2).

The same idea can be used for multidimensional Markov processes under $\widetilde{\mathbb{P}}$. A case of this was Example 1.3.2 of Section 1.3, in which the payoff of a derivative security was $V_3 = M_3 - S_3$, the difference between the stock price at time three and its maximum between times zero and three. Because only the stock price and its maximum-to-date appear in the payoff, we can use the two-dimensional Markov process $\{(S_n, M_n); n = 0, 1, 2, 3\}$ to treat this problem, which was done implicitly in that example.

Here we generalize Example 1.3.2 to an N-period binomial model with a derivative security whose payoff at time N is a function $v_N(S_N, M_N)$ of the stock price and the maximum stock price. (We do not mean that v_N is necessarily a function of *both* S_N and M_N but rather that these are the *only* random variables on which V_N depends. For example, we could have $V_N = (M_N - K)^+$. Even though the stock price does not appear in this particular V_N, we would need it to execute the pricing algorithm (2.5.9) below because the maximum-to-date process is not Markov by itself.) According to the "multi-step-head" Markov property, for any n between zero and N, there is a (nonrandom) function $v_n(s, m)$ such that the price of the option at time n is

$$V_n = v_n(S_n, M_n) = \widetilde{\mathbb{E}}_n\left[\frac{v_N(S_N, M_N)}{(1+r)^{N-n}}\right].$$

We can use the Independence Lemma 2.5.3 to derive an algorithm for computing the functions v_n. We always have the risk-neutral pricing formula (see (2.4.12))

$$V_n = \frac{1}{1+r}\widetilde{\mathbb{E}}_n[V_{n+1}]$$

relating the price of a derivative security at time n to its price at time $n+1$. Suppose that for some n between zero and $N-1$, we have computed the function v_{n+1} such that $V_{n+1} = v_{n+1}(S_{n+1}, M_{n+1})$. Then

$$\begin{aligned} V_n &= \frac{1}{1+r}\widetilde{\mathbb{E}}_n[V_{n+1}] \\ &= \frac{1}{1+r}\widetilde{\mathbb{E}}_n\left[v_{n+1}(S_{n+1}, M_{n+1})\right] \\ &= \frac{1}{1+r}\widetilde{\mathbb{E}}_n\left[v_{n+1}\left(S_n \cdot \frac{S_{n+1}}{S_n}, M_n \vee \left(S_n \cdot \frac{S_{n+1}}{S_n}\right)\right)\right]. \end{aligned}$$

To compute this last expression, we replace S_n and M_n by dummy variables s and m because they depend only on the first n tosses. We then take the unconditional expectation of $\frac{S_{n+1}}{S_n}$ because it does not depend on the first n tosses, i.e., we define

$$v_n(s,m) = \frac{1}{1+r}\widetilde{\mathbb{E}}_n\left[v_{n+1}\left(s\cdot\frac{S_{n+1}}{S_n}, m \vee \left(s\cdot\frac{S_{n+1}}{S_n}\right)\right)\right] \tag{2.5.8}$$
$$= \frac{1}{1+r}\Big[\tilde{p}v_{n+1}(us, m\vee(us)) + \tilde{q}v_{n+1}(ds, m\vee(ds))\Big].$$

The Independence Lemma 2.5.3 asserts that $V_n = v_n(S_n, M_n)$.

We will only need to know the value of $v_n(s,m)$ when $m \geq s$ since $M_n \geq S_n$. We can impose this condition in (2.5.8). But when $m \geq s$, if $d \leq 1$ as it usually is, we have $m \vee (ds) = m$. Therefore, we can rewrite (2.5.8) as

$$v_n(s,m) = \frac{1}{1+r}\Big[\tilde{p}v_{n+1}(us, m\vee(us)) + \tilde{q}v_{n+1}(ds, m)\Big], \tag{2.5.9}$$
$$m \geq s > 0,\ n = N-1, N-2, \ldots, 0.$$

This algorithm works for any derivative security whose payoff at time N depends only on the random variables S_N and M_N.

In Example 1.3.2, we were given that $V_3 = v_3(s,m)$, where $v_3(s,m) = m - s$. We used (2.5.9) to compute v_2, then used it again to compute v_1, and finally used it to compute v_0. These steps were carried out in Example 1.3.2.

In continuous time, we shall see that the analogue of recursive equations (2.5.9) are partial differential equations. The process that gets us from the continuous-time analogue of the risk-neutral pricing formula to these partial differential equations is the *Feynman-Kac Theorem.*

We summarize this discussion with a theorem.

Theorem 2.5.8. *Let $X_0, X_1, \ldots, X_N$ be a Markov process under the risk-neutral probability measure $\widetilde{\mathbb{P}}$ in the binomial model. Let $v_N(x)$ be a function of the dummy variable x, and consider a derivative security whose payoff at time N is $v_N(X_N)$. Then, for each n between 0 and N, the price V_n of this derivative security is some function v_n of X_n, i.e.,*

$$V_n = v_n(X_n),\ n = 0, 1, \ldots, N. \tag{2.5.10}$$

There is a recursive algorithm for computing v_n whose exact formula depends on the underlying Markov process $X_0, X_1, \ldots, X_N$. Analogous results hold if the underlying Markov process is multidimensional.

2.6 Summary

This chapter sets out the view of probability that begins with a random experiment having outcome ω. The collection of all possible outcomes is called the *sample space* Ω, and on this space we have a probability measure $\mathbb{P}$. When Ω is finite, we describe $\mathbb{P}$ by specifying for each $\omega \in \Omega$ the probability $\mathbb{P}(\omega)$ assigned to ω by $\mathbb{P}$. A random variable is a function X from Ω to $\mathbb{R}$, and the expectation of the random variable X is $\mathbb{E}X = \sum_{\omega\in\Omega} X(\omega)\mathbb{P}(\omega)$. If we

have a second probability measure $\widetilde{\mathbb{P}}$ on Ω, then we will have another way of computing the expectation, namely $\widetilde{\mathbb{E}}X = \sum_{\omega\in\Omega} X(\omega)\widetilde{\mathbb{P}}(\omega)$. The random variable X is the same in both cases, even though the two expectations are different. The point is that the random variable should not be thought of as a distribution. When we change probability measures, distributions (and hence expectations) will change, but random variables will not.

In the binomial model, we may see coin tosses $\omega_1 \ldots \omega_n$ and, based on this information, compute the conditional expectation of a random variable X that depends on coin tosses $\omega_1 \ldots \omega_n\omega_{n+1} \ldots \omega_N$. This is done by averaging over the possible outcomes of the "remaining" coin tosses $\omega_{n+1} \ldots \omega_N$. If we are computing the conditional expectation under the risk-neutral probabilities, this results in the formula

$$\widetilde{\mathbb{E}}_n[X](\omega_1 \ldots \omega_n) \tag{2.3.6}$$
$$= \sum_{\omega_{n+1}\ldots\omega_N} \tilde{p}^{\#H(\omega_{n+1}\ldots\omega_N)}\tilde{q}^{\#T(\omega_{n+1}\ldots\omega_N)}X(\omega_1 \ldots \omega_n\omega_{n+1} \ldots \omega_N).$$

This conditional expectation is a random variable because it depends on the first n coin tosses $\omega_1 \ldots \omega_n$. Conditional expectations have five fundamental properties, which are provided in Theorem 2.3.2.

In a multiperiod binomial model, a *martingale* under the risk-neutral probability measure $\widetilde{\mathbb{P}}$ is a sequence of random variables $M_0, M_1, \ldots, M_N$, where each M_n depends on only the first n coin tosses, and

$$M_n(\omega_1 \ldots \omega_n) = \widetilde{\mathbb{E}}_n[M_{n+1}](\omega_1 \ldots \omega_n)$$

no matter what the value of n and no matter what the coin tosses $\omega_1 \ldots \omega_n$ are. A martingale has no tendency to rise or fall. Conditioned on the information we have at time n, the expected value of the martingale at time $n+1$ is its value at time n.

Under the risk-neutral probability measure, the discounted stock price is a martingale, as is the discounted value of any portfolio that trades in the stock and money markets account. In particular, if X_n is the value of a portfolio at time n, then

$$\frac{X_n}{(1+r)^n} = \widetilde{\mathbb{E}}_n\left[\frac{X_N}{(1+r)^N}\right],\ 0 \leq n \leq N.$$

If we want to have X_N agree with the value V_N of a derivative security at its expiration time N, then we must have

$$\frac{X_n}{(1+r)^n} = \widetilde{\mathbb{E}}_n\left[\frac{X_N}{(1+r)^N}\right] = \widetilde{\mathbb{E}}_n\left[\frac{V_N}{(1+r)^N}\right] \tag{2.4.9}$$

at all times $n = 0, 1, \ldots, N$. When a portfolio does this, we define the value V_n of the derivative security at time n to be X_n, and we thus have the risk-neutral pricing formula

$$V_n = \widetilde{\mathbb{E}}_n \left[\frac{V_N}{(1+r)^{N-n}} \right]. \tag{2.4.11}$$

A *Markov process* is a sequence of random variables $X_0, X_1, \dots, X_N$ with the following property. Suppose n is a time between 0 and $N-1$, we have observed the first n coin tosses $\omega_1 \dots \omega_n$, and we want to estimate either a function of X_{n+1} or, more generally, a function of X_{n+k} for some k between 1 and $N-n$. We know both the individual coin tosses $\omega_1 \dots \omega_n$ and the resulting value $X_n(\omega_1 \dots \omega_n)$ and can base our estimate on this information. For a Markov process, knowledge of the individual coin tosses (the "path") does not provide any information relevant to this estimation problem beyond that information already contained in our knowledge of the value $X_n(\omega_1 \dots \omega_n)$.

Consider an underlying asset-price process $X_0, X_1, \dots, X_N$ that is Markov under the risk-neutral measure and a derivative security payoff at time N that is a function of this asset price at time N; i.e., $V_N = v_N(X_N)$. The price of the derivative security at all times n prior to expiration is a function of the underlying asset price at those times; i.e.,

$$V_n = v_n(X_n), \; n = 0, 1, \dots, N. \tag{2.4.11}$$

In this notation, V_n is a random variable depending on the coin tosses $\omega_1 \dots \omega_n$. It is potentially path-dependent. On the other hand, $v_n(x)$ is a function of a real number x. When we replace x by the random variable X_n, then $v_n(X_n)$ also becomes random, but in a way that is guaranteed not to be path-dependent. Equation (2.4.11) thus guarantees that the price of the derivative security is not path-dependent.

2.7 Notes

The sample space view of probability theory dates back to Kolmogorov [29], who developed it in a way that extends to infinite probability spaces. We take up this subject in Chapters 1 and 2 of Volume II. Martingales were invented by Doob [13], who attributes the idea and the name "martingale" to a gambling strategy discussed by Ville [43].

The risk-neutral pricing formula is due to Harrison and Kreps [17] and Harrison and Pliska [18].

2.8 Exercises

Exercise 2.1. Using Definition 2.1.1, show the following.

(i) If A is an event and A^c denotes its complement, then $\mathbb{P}(A^c) = 1 - \mathbb{P}(A)$.

(ii) If $A_1, A_2, \dots, A_N$ is a finite set of events, then

$$\mathbb{P}\left(\bigcup_{n=1}^{N} A_n \right) \leq \sum_{n=1}^{N} \mathbb{P}(A_n). \tag{2.8.1}$$

If the events $A_1, A_2, \dots, A_N$ are disjoint, then equality holds in (2.8.1).

Exercise 2.2. Consider the stock price S_3 in Figure 2.3.1.

(i) What is the distribution of S_3 under the risk-neutral probabilities $\tilde{p} = \frac{1}{2}$, $\tilde{q} = \frac{1}{2}$.

(ii) Compute $\widetilde{\mathbb{E}}S_1$, $\widetilde{\mathbb{E}}S_2$, and $\widetilde{\mathbb{E}}S_3$. What is the average rate of growth of the stock price under $\widetilde{\mathbb{P}}$?

(iii) Answer (i) and (ii) again under the actual probabilities $p = \frac{2}{3}$, $q = \frac{1}{3}$.

Exercise 2.3. Show that a convex function of a martingale is a submartingale. In other words, let $M_0, M_1, \dots, M_N$ be a martingale and let φ be a convex function. Show that $\varphi(M_0), \varphi(M_1), \dots, \varphi(M_N)$ is a submartingale.

Exercise 2.4. Toss a coin repeatedly. Assume the probability of head on each toss is $\frac{1}{2}$, as is the probability of tail. Let $X_j = 1$ if the jth toss results in a head and $X_j = -1$ if the jth toss results in a tail. Consider the stochastic process $M_0, M_1, M_2, \dots$ defined by $M_0 = 0$ and

$$M_n = \sum_{j=1}^{n} X_j, \; n \geq 1.$$

This is called a *symmetric random walk*; with each head, it steps up one, and with each tail, it steps down one.

(i) Using the properties of Theorem 2.3.2, show that $M_0, M_1, M_2, \dots$ is a martingale.

(ii) Let σ be a positive constant and, for $n \geq 0$, define

$$S_n = e^{\sigma M_n} \left(\frac{2}{e^{\sigma} + e^{-\sigma}} \right)^n.$$

Show that $S_0, S_1, S_2, \dots$ is a martingale. Note that even though the symmetric random walk M_n has no tendency to grow, the "geometric symmetric random walk" $e^{\sigma M_n}$ does have a tendency to grow. This is the result of putting a martingale into the (convex) exponential function (see Exercise 2.3). In order to again have a martingale, we must "discount" the geometric symmetric random walk, using the term $\frac{2}{e^{\sigma}+e^{-\sigma}}$ as the discount rate. This term is strictly less than one unless $\sigma = 0$.

Exercise 2.5. Let $M_0, M_1, M_2, \dots$ be the symmetric random walk of Exercise 2.4, and define $I_0 = 0$ and

$$I_n = \sum_{j=0}^{n-1} M_j (M_{j+1} - M_j), \; n = 1, 2, \dots.$$

(i) Show that

$$I_n = \frac{1}{2}M_n^2 - \frac{n}{2}.$$

(ii) Let n be an arbitrary nonnegative integer, and let $f(i)$ be an arbitrary function of a variable i. In terms of n and f, define another function $g(i)$ satisfying

$$E_n\big[f(I_{n+1})\big] = g(I_n).$$

Note that although the function $g(I_n)$ on the right-hand side of this equation may depend on n, the only random variable that may appear in its argument is I_n; the random variable M_n may not appear. You will need to use the formula in part (i). The conclusion of part (ii) is that the process $I_0, I_1, I_2, \ldots$ is a Markov process.

Exercise 2.6 (Discrete-time stochastic integral). Suppose $M_0, M_1, \ldots, M_N$ is a martingale, and let $\Delta_0, \Delta_1, \ldots, \Delta_{N-1}$ be an adapted process. Define the *discrete-time stochastic integral* (sometimes called a *martingale transform*) $I_0, I_1, \ldots, I_N$ by setting $I_0 = 0$ and

$$I_n = \sum_{j=0}^{n-1} \Delta_j (M_{j+1} - M_j),\ n = 1, \ldots, N.$$

Show that $I_0, I_1, \ldots, I_N$ is a martingale.

Exercise 2.7. In a binomial model, give an example of a stochastic process that is a martingale but is not Markov.

Exercise 2.8. Consider an N-period binomial model.

(i) Let $M_0, M_1, \ldots, M_N$ and $M_0', M_1', \ldots, M_N'$ be martingales under the risk-neutral measure $\widetilde{\mathbb{P}}$. Show that if $M_N = M_N'$ (for every possible outcome of the sequence of coin tosses), then, for each n between 0 and N, we have $M_n = M_n'$ (for every possible outcome of the sequence of coin tosses).

(ii) Let V_N be the payoff at time N of some derivative security. This is a random variable that can depend on all N coin tosses. Define recursively $V_{N-1}, V_{N-2}, \ldots, V_0$ by the algorithm (1.2.16) of Chapter 1. Show that

$$V_0, \frac{V_1}{1+r}, \ldots, \frac{V_{N-1}}{(1+r)^{N-1}}, \frac{V_N}{(1+r)^N}$$

is a martingale under $\widetilde{\mathbb{P}}$.

(iii) Using the risk-neutral pricing formula (2.4.11) of this chapter, define

$$V_n' = \widetilde{\mathbb{E}}_n\left[\frac{V_N}{(1+r)^{N-n}}\right],\ n = 0, 1, \ldots, N-1.$$

Show that

$$V_0', \frac{V_1'}{1+r}, \ldots, \frac{V_{N-1}'}{(1+r)^{N-1}}, \frac{V_N}{(1+r)^N}$$

is a martingale.

(iv) Conclude that $V_n = V_n'$ for every n (i.e., the algorithm (1.2.16) of Theorem 1.2.2 of Chapter 1 gives the same derivative security prices as the risk-neutral pricing formula (2.4.11) of Chapter 2).

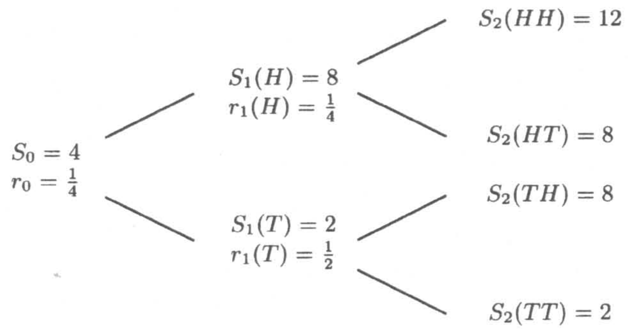

Fig. 2.8.1. A stochastic volatility, random interest rate model.

Exercise 2.9 (Stochastic volatility, random interest rate). Consider a two-period stochastic volatility, random interest rate model of the type described in Exercise 1.9 of Chapter 1. The stock prices and interest rates are shown in Figure 2.8.1.

(i) Determine risk-neutral probabilities

$$\widetilde{\mathbb{P}}(HH),\ \widetilde{\mathbb{P}}(HT),\ \widetilde{\mathbb{P}}(TH),\ \widetilde{\mathbb{P}}(TT),$$

such that the time-zero value of an option that pays off V_2 at time two is given by the risk-neutral pricing formula

$$V_0 = \widetilde{\mathbb{E}}\left[\frac{V_2}{(1+r_0)(1+r_1)}\right].$$

(ii) Let $V_2 = (S_2 - 7)^+$. Compute V_0, $V_1(H)$, and $V_1(T)$.

(iii) Suppose an agent sells the option in (ii) for V_0 at time zero. Compute the position Δ_0 she should take in the stock at time zero so that at time one, regardless of whether the first coin toss results in head or tail, the value of her portfolio is V_1.

(iv) Suppose in (iii) that the first coin toss results in head. What position $\Delta_1(H)$ should the agent now take in the stock to be sure that, regardless of whether the second coin toss results in head or tail, the value of her portfolio at time two will be $(S_2 - 7)^+$?

Exercise 2.10 (Dividend-paying stock). We consider a binomial asset pricing model as in Chapter 1, except that, after each movement in the stock price, a dividend is paid and the stock price is reduced accordingly. To describe this in equations, we define

$$Y_{n+1}(\omega_1 \dots \omega_n \omega_{n+1}) = \begin{cases} u, \text{ if } \omega_{n+1} = H, \\ d, \text{ if } \omega_{n+1} = T. \end{cases}$$

Note that Y_{n+1} depends only on the $(n+1)$st coin toss. In the binomial model of Chapter 1, $Y_{n+1}S_n$ was the stock price at time $n+1$. In the dividend-paying model considered here, we have a random variable $A_{n+1}(\omega_1 \dots \omega_n \omega_{n+1})$, taking values in $(0,1)$, and the dividend paid at time $n+1$ is $A_{n+1}Y_{n+1}S_n$. After the dividend is paid, the stock price at time $n+1$ is

$$S_{n+1} = (1 - A_{n+1})Y_{n+1}S_n.$$

An agent who begins with initial capital X_0 and at each time n takes a position of Δ_n shares of stock, where Δ_n depends only on the first n coin tosses, has a portfolio value governed by the wealth equation (see (2.4.6))

$$\begin{aligned} X_{n+1} &= \Delta_n S_{n+1} + (1+r)(X_n - \Delta_n S_n) + \Delta_n A_{n+1} Y_{n+1} S_n \\ &= \Delta_n Y_{n+1} S_n + (1+r)(X_n - \Delta_n S_n). \end{aligned} \tag{2.8.2}$$

(i) Show that the discounted wealth process is a martingale under the risk-neutral measure (i.e., Theorem 2.4.5 still holds for the wealth process (2.8.2)). As usual, the risk-neutral measure is still defined by the equations

$$\tilde{p} = \frac{1+r-d}{u-d}, \quad \tilde{q} = \frac{u-1-r}{u-d}.$$

(ii) Show that the risk-neutral pricing formula still applies (i.e., Theorem 2.4.7 holds for the dividend-paying model).

(iii) Show that the discounted stock price is not a martingale under the risk-neutral measure (i.e., Theorem 2.4.4 no longer holds). However, if A_{n+1} is a constant $a \in (0,1)$, regardless of the value of n and the outcome of the coin tossing $\omega_1 \dots \omega_{n+1}$, then $\frac{S_n}{(1-a)^n(1+r)^n}$ is a martingale under the risk-neutral measure.

Exercise 2.11 (Put–call parity). Consider a stock that pays no dividend in an N-period binomial model. A European call has payoff $C_N = (S_N - K)^+$ at time N. The price C_n of this call at earlier times is given by the risk-neutral pricing formula (2.4.11):

$$C_n = \widetilde{\mathbb{E}}_n \left[\frac{C_N}{(1+r)^{N-n}} \right], \quad n = 0, 1, \dots, N-1.$$

Consider also a put with payoff $P_N = (K - S_N)^+$ at time N, whose price at earlier times is

$$P_n = \widetilde{\mathbb{E}}_n \left[\frac{P_N}{(1+r)^{N-n}} \right], \quad n = 0, 1, \dots, N-1.$$

Finally, consider a *forward contract* to buy one share of stock at time N for K dollars. The price of this contract at time N is $F_N = S_N - K$, and its price at earlier times is

$$F_n = \widetilde{\mathbb{E}}_n \left[\frac{F_N}{(1+r)^{N-n}} \right], \quad n = 0, 1, \dots, N-1.$$

(Note that, unlike the call, the forward contract requires that the stock be purchased at time N for K dollars and has a negative payoff if $S_N < K$.)

(i) If at time zero you buy a forward contract and a put, and hold them until expiration, explain why the payoff you receive is the same as the payoff of a call; i.e., explain why $C_N = F_N + P_N$.

(ii) Using the risk-neutral pricing formulas given above for C_n, P_n, and F_n and the linearity of conditional expectations, show that $C_n = F_n + P_n$ for every n.

(iii) Using the fact that the discounted stock price is a martingale under the risk-neutral measure, show that $F_0 = S_0 - \frac{K}{(1+r)^N}$.

(iv) Suppose you begin at time zero with F_0, buy one share of stock, borrowing money as necessary to do that, and make no further trades. Show that at time N you have a portfolio valued at F_N. (This is called a *static replication* of the forward contract. If you sell the forward contract for F_0 at time zero, you can use this static replication to hedge your short position in the forward contract.)

(v) The *forward price* of the stock at time zero is defined to be that value of K that causes the forward contract to have price zero at time zero. The forward price in this model is $(1+r)^N S_0$. Show that, at time zero, the price of a call struck at the forward price is the same as the price of a put struck at the forward price. This fact is called *put–call parity.*

(vi) If we choose $K = (1+r)^N S_0$, we just saw in (v) that $C_0 = P_0$. Do we have $C_n = P_n$ for every n?

Exercise 2.12 (Chooser option). Let $1 \leq m \leq N-1$ and $K > 0$ be given. A *chooser option* is a contract sold at time zero that confers on its owner the right to receive either a call or a put at time m. The owner of the chooser may wait until time m before choosing. The call or put chosen expires at time N with strike price K. Show that the time-zero price of a chooser option is the sum of the time-zero price of a put, expiring at time N and having strike price K, and a call, expiring at time m and having strike price $\frac{K}{(1+r)^{N-m}}$. (Hint: Use put–call parity (Exercise 2.11).)

Exercise 2.13 (Asian option). Consider an N-period binomial model. An *Asian option* has a payoff based on the average stock price, i.e.,

$$V_N = f\left(\frac{1}{N+1}\sum_{n=0}^{N} S_n\right),$$

where the function f is determined by the contractual details of the option.

(i) Define $Y_n = \sum_{k=0}^{n} S_k$ and use the Independence Lemma 2.5.3 to show that the two-dimensional process (S_n, Y_n), $n = 0, 1, \ldots, N$ is Markov.
(ii) According to Theorem 2.5.8, the price V_n of the Asian option at time n is some function v_n of S_n and Y_n; i.e.,

$$V_n = v_n(S_n, Y_n), \quad n = 0, 1, \ldots, N.$$

Give a formula for $v_N(s, y)$, and provide an algorithm for computing $v_n(s, y)$ in terms of v_{n+1}.

Exercise 2.14 (Asian option continued). Consider an N-period binomial model, and let M be a fixed number between 0 and $N-1$. Consider an Asian option whose payoff at time N is

$$V_N = f\left(\frac{1}{N-M}\sum_{n=M+1}^{N} S_n\right),$$

where again the function f is determined by the contractual details of the option.

(i) Define

$$Y_n = \begin{cases} 0, & \text{if } 0 \le n \le M, \\ \sum_{k=M+1}^{n} S_k, & \text{if } M+1 \le n \le N. \end{cases}$$

Show that the two-dimensional process (S_n, Y_n), $n = 0, 1, \ldots, N$ is Markov (under the risk-neutral measure $\widetilde{\mathbb{P}}$).
(ii) According to Theorem 2.5.8, the price V_n of the Asian option at time n is some function v_n of S_n and Y_n, i.e.,

$$V_n = v_n(S_n, Y_n), \quad n = 0, 1, \ldots, N.$$

Of course, when $n \le M$, Y_n is not random and does not need to be included in this function. Thus, for such n we should seek a function v_n of S_n alone and have

$$V_n = \begin{cases} v_n(S_n), & \text{if } 0 \le n \le M, \\ v_n(S_n, Y_n), & \text{if } M+1 \le n \le N. \end{cases}$$

Give a formula for $v_N(s, y)$, and provide an algorithm for computing v_n in terms of v_{n+1}. Note that the algorithm is different for $n < M$ and $n > M$, and there is a separate transition formula for $v_M(s)$ in terms of $v_{M+1}(\cdot, \cdot)$.

3

State Prices

3.1 Change of Measure

In the binomial no-arbitrage pricing model of Chapter 1 and also in the continuous-time models formulated in Chapters 4 and 5 of Volume II, there are two probability measures that merit our attention. One is the *actual probability measure*, by which we mean the one that we seek by empirical estimation of the model parameters. The other is the *risk-neutral probability measure*, under which the discounted prices of assets are martingales. These two probability measures give different weights to the asset-price paths in the model. They agree, however, on which price paths are possible (i.e., which paths have positive probability of occurring); they disagree only on what these positive probabilities are. The actual probabilities are the "right" ones. The risk-neutral probabilities are a fictitious but helpful construct because they allow us to neatly summarize the result of solving systems of equations (see, e.g., the system (1.1.3), (1.1.4) of Chapter 1, which leads to the formula (1.1.7) of that chapter).

Let us more generally consider a finite sample space Ω on which we have two probability measures $\mathbb{P}$ and $\widetilde{\mathbb{P}}$. Let us assume that $\mathbb{P}$ and $\widetilde{\mathbb{P}}$ both give positive probability to every element of Ω, so we can form the quotient

$$Z(\omega) = \frac{\widetilde{\mathbb{P}}(\omega)}{\mathbb{P}(\omega)}. \tag{3.1.1}$$

Because it depends on the outcome ω of a random experiment, Z is a random variable. It is called the *Radon-Nikodým derivative of $\widetilde{\mathbb{P}}$ with respect to $\mathbb{P}$*, although in this context of a finite sample space Ω, it is really a quotient rather than a derivative. The random variable Z has three important properties, which we state as a theorem.

Theorem 3.1.1. *Let $\mathbb{P}$ and $\widetilde{\mathbb{P}}$ be probability measures on a finite sample space Ω, assume that $\mathbb{P}(\omega) > 0$ and $\widetilde{\mathbb{P}}(\omega) > 0$ for every $\omega \in \Omega$, and define the random variable Z by (3.1.1). Then we have the following:*

(i) $\mathbb{P}(Z > 0) = 1$;
(ii) $\mathbb{E}Z = 1$;
(iii) for any random variable Y,

$$\widetilde{\mathbb{E}}Y = \mathbb{E}[ZY]. \tag{3.1.2}$$

PROOF: Property (i) follows immediately from the fact that we have assumed $\widetilde{\mathbb{P}}(\omega) > 0$ for every $\omega \in \Omega$. Property (ii) can be verified by the computation

$$\mathbb{E}Z = \sum_{\omega\in\Omega} Z(\omega)\mathbb{P}(\omega) = \sum_{\omega\in\Omega} \frac{\widetilde{\mathbb{P}}(\omega)}{\mathbb{P}(\omega)}\mathbb{P}(\omega) = \sum_{\omega\in\Omega} \widetilde{\mathbb{P}}(\omega) = 1,$$

the last equality following from the fact that $\widetilde{\mathbb{P}}$ is a probability measure. The following similar computation verifies property (iii):

$$\begin{aligned} \widetilde{\mathbb{E}}Y &= \sum_{\omega\in\Omega} Y(\omega)\widetilde{\mathbb{P}}(\omega) = \sum_{\omega\in\Omega} Y(\omega)\frac{\widetilde{\mathbb{P}}(\omega)}{\mathbb{P}(\omega)}\mathbb{P}(\omega) \\ &= \sum_{\omega\in\Omega} Y(\omega)Z(\omega)\mathbb{P}(\omega) = \mathbb{E}[ZY]. \end{aligned}$$ □

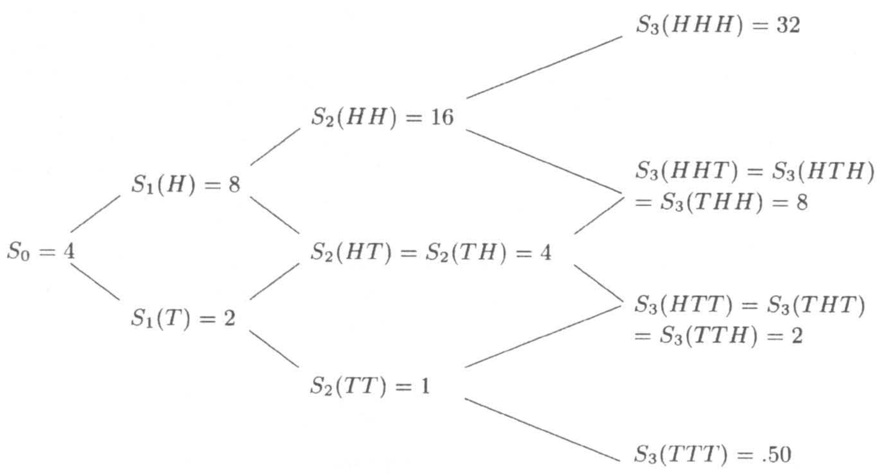

Fig. 3.1.1. A three-period model.

Example 3.1.2. Consider again the three-period model of Figure 3.1.1. The underlying probability space is

$$\Omega = \{HHH, HHT, HTH, HTT, THH, THT, TTH, TTT\}.$$

We take $p = \frac{2}{3}$ as the actual probability of a head and $q = \frac{1}{3}$ as the actual probability of a tail. Then the actual probability measure is

$$\begin{gathered} \mathbb{P}(HHH) = \frac{8}{27}, \quad \mathbb{P}(HHT) = \frac{4}{27}, \quad \mathbb{P}(HTH) = \frac{4}{27}, \quad \mathbb{P}(HTT) = \frac{2}{27}, \\ \mathbb{P}(THH) = \frac{4}{27}, \quad \mathbb{P}(THT) = \frac{2}{27}, \quad \mathbb{P}(TTH) = \frac{2}{27}, \quad \mathbb{P}(TTT) = \frac{1}{27}. \end{gathered} \tag{3.1.3}$$

We take the interest rate to be $r = \frac{1}{4}$, and then the risk-neutral probability of a head is $\tilde{p} = \frac{1}{2}$ and the risk-neutral probability of a tail is $\tilde{q} = \frac{1}{2}$. The risk-neutral probability measure is

$$\begin{gathered} \widetilde{\mathbb{P}}(HHH) = \frac{1}{8}, \quad \widetilde{\mathbb{P}}(HHT) = \frac{1}{8}, \quad \widetilde{\mathbb{P}}(HTH) = \frac{1}{8}, \quad \widetilde{\mathbb{P}}(HTT) = \frac{1}{8}, \\ \widetilde{\mathbb{P}}(THH) = \frac{1}{8}, \quad \widetilde{\mathbb{P}}(THT) = \frac{1}{8}, \quad \widetilde{\mathbb{P}}(TTH) = \frac{1}{8}, \quad \widetilde{\mathbb{P}}(TTT) = \frac{1}{8}. \end{gathered} \tag{3.1.4}$$

Therefore, the Radon-Nikodým derivative of $\widetilde{\mathbb{P}}$ with respect to $\mathbb{P}$ is

$$\begin{gathered} Z(HHH) = \frac{27}{64}, \quad Z(HHT) = \frac{27}{32}, \quad Z(HTH) = \frac{27}{32}, \quad Z(HTT) = \frac{27}{16}, \\ Z(THH) = \frac{27}{32}, \quad Z(THT) = \frac{27}{16}, \quad Z(TTH) = \frac{27}{16}, \quad Z(TTT) = \frac{27}{8}. \end{gathered} \tag{3.1.5}$$

In Example 1.2.4 of Chapter 1, for this model we determined the time-zero price of a lookback option whose payoff at time three was given by

$$\begin{gathered} V_3(HHH) = 0, \quad V_3(HHT) = 8, \quad V_3(HTH) = 0, \quad V_3(HTT) = 6, \\ V_3(THH) = 0, \quad V_3(THT) = 2, \quad V_3(TTH) = 2, \quad V_3(TTT) = 3.50. \end{gathered}$$

According to the risk-neutral pricing formula (2.4.11) of Chapter 2, this time-zero value is

$$\begin{aligned} V_0 &= \widetilde{\mathbb{E}}\frac{V_3}{(1+r)^3} \\ &= \left(\frac{4}{5}\right)^3 \sum_{\omega\in\Omega} V_3(\omega)\widetilde{\mathbb{P}}(\omega) \\ &= 0.512\left[0\cdot\frac{1}{8} + 8\cdot\frac{1}{8} + 0\cdot\frac{1}{8} + 6\cdot\frac{1}{8} + 0\cdot\frac{1}{8} + 2\cdot\frac{1}{8} + 2\cdot\frac{1}{8} + 3.50\cdot\frac{1}{8}\right] \\ &= 1.376, \end{aligned} \tag{3.1.6}$$

which is the number determined in Example 1.2.4 of Chapter 1 to be the cost at time zero of setting up a replicating portfolio. Using the random variable Z, we can rewrite (3.1.6) as

$$
\begin{aligned}
V_0 &= \mathbb{E}\frac{V_3 Z}{(1+r)^3}\\
&= \left(\frac{4}{5}\right)^3 \sum_{\omega\in\Omega} V_3(\omega)Z(\omega)\mathbb{P}(\omega)\\
&= 0.512\Big[0\cdot\frac{27}{64}\cdot\frac{8}{27}+8\cdot\frac{27}{32}\cdot\frac{4}{27}+0\cdot\frac{27}{32}\cdot\frac{4}{27}+6\cdot\frac{27}{16}\cdot\frac{2}{27}\\
&\qquad +0\cdot\frac{27}{32}\cdot\frac{4}{27}+2\cdot\frac{27}{16}\cdot\frac{2}{27}+2\cdot\frac{27}{16}\cdot\frac{2}{27}+3.50\cdot\frac{27}{8}\cdot\frac{1}{27}\Big]\\
&= 1.376, \qquad (3.1.7)
\end{aligned}
$$

The advantage of (3.1.7) over (3.1.6) is that (3.1.7) makes no reference to the risk-neutral measure. However, it does not simply compute the expected discounted payoff of the option under the actual probability measure but rather first weights these payoffs using the random variable Z. This leads to the idea of *state prices*, which we formalize in the next definition. □

Definition 3.1.3. *In the N-period binomial model with actual probability measure $\mathbb{P}$ and risk-neutral probability measure $\widetilde{\mathbb{P}}$, let Z denote the Radon-Nikodým derivative of $\widetilde{\mathbb{P}}$ with respect to $\mathbb{P}$; i.e.,*

$$
Z(\omega_1\ldots\omega_N) = \frac{\widetilde{\mathbb{P}}(\omega_1\ldots\omega_N)}{\mathbb{P}(\omega_1\ldots\omega_N)} = \left(\frac{\tilde{p}}{p}\right)^{\#H(\omega_1\ldots\omega_N)}\left(\frac{\tilde{q}}{q}\right)^{\#T(\omega_1\ldots\omega_N)}, \qquad (3.1.8)
$$

where $\#H(\omega_1\ldots\omega_N)$ denotes the number of heads appearing in the sequence $\omega_1\ldots\omega_N$ and $\#T(\omega_1\ldots\omega_N)$ denotes the number of tails appearing in this sequence. The state price density *random variable is*

$$
\zeta(\omega) = \frac{Z(\omega)}{(1+r)^N}, \qquad (3.1.9)
$$

and $\zeta(\omega)\mathbb{P}(\omega)$ is called the state price *corresponding to ω.*

Let $\overline{\omega} = \overline{\omega}_1\ldots\overline{\omega}_N$ be a particular coin toss sequence in the N-period model, and consider a derivative security that pays off 1 if $\overline{\omega}$ occurs and otherwise pays off 0; i.e.,

$$
V_N(\omega) = \begin{cases} 1, & \text{if } \omega = \overline{\omega},\\ 0, & \text{otherwise.}\end{cases}
$$

According to the risk-neutral pricing formula, the value of this derivative security at time zero is

$$
\widetilde{\mathbb{E}}\frac{V_N}{(1+r)^N} = \frac{\widetilde{\mathbb{P}}(\overline{\omega})}{(1+r)^N} = \frac{Z(\overline{\omega})\mathbb{P}(\overline{\omega})}{(1+r)^N} = \zeta(\overline{\omega})\mathbb{P}(\overline{\omega}).
$$

We see that the state price $\zeta(\overline{\omega})\mathbb{P}(\overline{\omega})$ tells the price at time zero of a contract that pays 1 at time N if and only if $\overline{\omega}$ occurs. This price should include a discount from time N to time zero to account for the time value of money, and the term $\frac{1}{(1+r)^N}$ does indeed appear in (3.1.9). It is natural to expect the price to take into account the probability that $\overline{\omega}$ will occur, and therefore we have arranged the formulas so that $\mathbb{P}(\overline{\omega})$ is one of the factors in the state price. However, these two factors alone cannot tell the whole story because they do not account for risk. If we were to use these terms alone, and take the time-zero price of a derivative security to be $\mathbb{E}\frac{V_N}{(1+r)^N}$, then the time-zero price of an asset would depend only on its expected return under the actual probability measure. In fact, the price of an asset depends on both its expected return and the risk it presents. The remaining term appearing in the state price corresponding to $\overline{\omega}$, $Z(\overline{\omega})$, accounts for risk. For example, in (3.1.5) we see that Z discounts the importance of the stock price paths that end above the initial stock price $S_0 = 4$ because $Z < 1$ whenever there are two or three heads in the three coin tosses, but Z inflates the importance of the stock price paths that end below the initial stock price. The effect of this is to make holding the stock appear less favorable than one would infer from simply computing $\mathbb{E}\left[\left(\frac{4}{5}\right)^3 S_3\right]$, its discounted expected value at time three.

The state price $\zeta(\overline{\omega})\mathbb{P}(\overline{\omega})$ tells us the time-zero price of a contract that pays 1 at time N if and only if $\overline{\omega}$ occurs. The state price density $\frac{Z(\overline{\omega})}{(1+r)^N}$ tells us the time-zero price of this contract per unit of actual probability. For this reason, we call it a density.

Of course, most contracts make payoffs for several different values of ω, and these payoffs are not all necessarily 1. Such a contract can be regarded as a portfolio of simple contracts, each of which pays off 1 if and only if some particular ω occurs, and their prices can be computed by summing up the prices of these components. To see this, recall from (2.4.11) of Chapter 2 the risk-neutral pricing formula $V_0 = \widetilde{\mathbb{E}}\frac{V_N}{(1+r)^N}$ for the time-zero price of an arbitrary derivative security paying V_N at time N. In terms of the state price density, this can be rewritten simply as

$$
V_0 = \mathbb{E}[\zeta V_N] = \sum_{\omega\in\Omega} V_N(\omega)\zeta(\omega)\mathbb{P}(\omega). \qquad (3.1.10)
$$

Equation (3.1.7) is a special case of this, where the $\zeta(\omega)$ term is separated into its factors $\frac{1}{(1+r)^3}$ and $Z(\omega)$.

3.2 Radon-Nikodým Derivative Process

In the previous section, we considered the Radon-Nikodým derivative of the risk-neutral probability measure with respect to the actual probability measure in an N-period binomial model. This random variable Z depends on the N coin tosses in the model. To get related random variables that depend on fewer coin tosses, we can estimate Z based on the information at time $n < N$.

This procedure of estimation will occur in other contexts as well, and thus we give a general result that does not require that Z is a Radon-Nikodým derivative.

Theorem 3.2.1. *Let Z be a random variable in an N-period binomial model. Define*

$$Z_n = \mathbb{E}_n Z, \quad n = 0, 1, \ldots, N. \tag{3.2.1}$$

Then Z_n, $n = 0, 1, \ldots, N$, is a martingale under $\mathbb{P}$.

PROOF: For $n = 0, 1, \ldots, N-1$, we use the "iterated conditioning" property of Theorem 2.3.2(iii) of Chapter 2 to compute

$$\mathbb{E}_n\big[Z_{n+1}\big] = \mathbb{E}_n\big[\mathbb{E}_{n+1}[Z]\big] = \mathbb{E}_n\big[Z\big] = Z_n.$$

This shows that Z_n, $n = 0, 1, \ldots, N$ is a martingale. □

Remark 3.2.2. Although Theorem 3.2.1 is stated for the probability measure $\mathbb{P}$, the analogous theorem is true under the risk-neutral probability measure $\widetilde{\mathbb{P}}$. The proof is the same. □

When successive estimates of a random variable are made, the estimates become more precise with increasing time (and information). However, Theorem 3.2.1 says they have no tendency to rise or fall. If a later estimate were on average higher than an earlier estimate, this tendency to rise would have already been incorporated into the earlier estimate. This is similar to the situation with an efficient stock market. If a stock were known to outperform other stocks having the same level of risk, this fact would have already been incorporated into the current price of the stock and thereby raise it to the point where the superior performance was no longer possible.

Example 3.2.3. Consider the three-period model of Example 3.1.2. In that example, we determined the Radon-Nikodým derivative of $\widetilde{\mathbb{P}}$ with respect to $\mathbb{P}$ to be given by (3.1.5). For $n = 0, 1, 2, 3$, we define $Z_n = \mathbb{E}_n[Z]$. In particular, $Z_3(\omega_1\omega_2\omega_3) = Z(\omega_1\omega_2\omega_3)$ for all $\omega_1\omega_2\omega_3$. We compute

$$Z_2(HH) = \frac{2}{3}Z_3(HHH) + \frac{1}{3}Z_3(HHT) = \frac{9}{16},$$

$$Z_2(HT) = \frac{2}{3}Z_3(HTH) + \frac{1}{3}Z_2(HTT) = \frac{9}{8},$$

$$Z_2(TH) = \frac{2}{3}Z_3(THH) + \frac{1}{3}Z_2(THT) = \frac{9}{8},$$

$$Z_2(TT) = \frac{2}{3}Z_2(TTH) + \frac{1}{3}Z^2(TTT) = \frac{9}{4}.$$

According to its definition, $Z_1 = \mathbb{E}_1[Z]$, but Theorem 3.2.1 allows us to compute it using the martingale formula $Z_1 = \mathbb{E}_1[Z_2]$. This leads to the equations

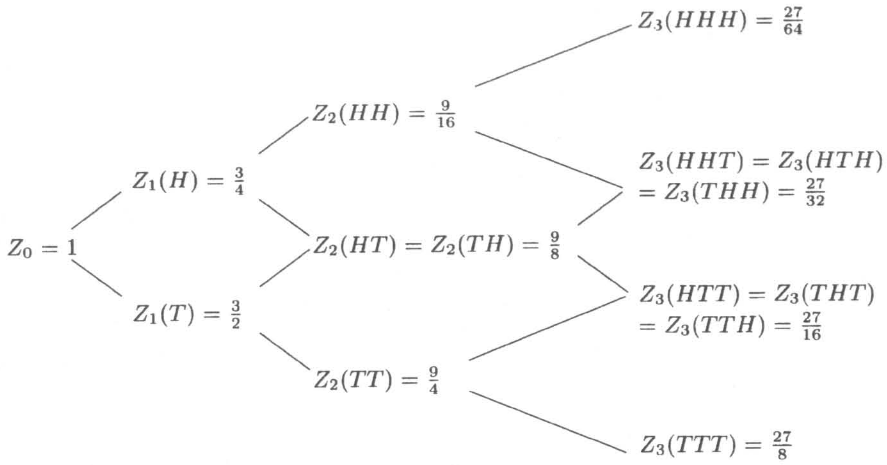

Fig. 3.2.1. A Radon-Nikodým derivative process.

$$Z_1(H) = \frac{2}{3}Z_2(HH) + \frac{1}{3}Z_2(HT) = \frac{3}{4},$$

$$Z_1(T) = \frac{2}{3}Z_1(TH) + \frac{1}{3}Z_1(TT) = \frac{3}{2}.$$

According to its definition, $Z_0 = \mathbb{E}Z$, which must be 1 because of Theorem 3.1.1(ii). We can also compute it using the martingale formula $Z_0 = \mathbb{E}_0[Z_1] = \mathbb{E}Z_1$, and this leads to

$$Z_0 = \tfrac{2}{3}Z_1(H) + \tfrac{1}{3}Z_1(T) = 1.$$

The process Z_n, $n = 0, 1, 2, 3$, is shown in Figure 3.2.1. □

Definition 3.2.4. *In an N-period binomial model, let $\mathbb{P}$ be the actual probability measure, $\widetilde{\mathbb{P}}$ the risk-neutral probability measure, and assume that $\mathbb{P}(\omega) > 0$ and $\widetilde{\mathbb{P}}(\omega) > 0$ for every sequence of coin tosses ω. Define the Radon-Nikodým derivative (random variable) $Z(\omega) = \frac{\widetilde{\mathbb{P}}(\omega)}{\mathbb{P}(\omega)}$ for every ω. The* Radon-Nikodým derivative process *is*

$$Z_n = \mathbb{E}_n[Z], \quad n = 0, 1, \ldots, N. \tag{3.2.2}$$

In particular, $Z_N = Z$ and $Z_0 = 1$.

In the context of Definition 3.2.4, we can compute the risk-neutral expectation of a random variable Y by computing under the actual probability measure the expectation $E[ZY]$. If Y only depends on the first n coin tosses, where $n < N$, this computation can be simplified further.

Lemma 3.2.5. *Assume the conditions of Definition 3.2.4. Let n be a positive integer between 0 and N, and let Y be a random variable depending only on the first n coin tosses. Then*

$$\widetilde{\mathbb{E}}Y = \mathbb{E}[Z_n Y]. \tag{3.2.3}$$

PROOF: We use Theorem 3.1.1(iii) of this chapter, Theorem 2.3.2(iii) of Chapter 2 (iterated conditioning), Theorem 2.3.2(ii) of Chapter 2 (taking out what is known), and the definition of Z_n, in that order, to justify each of the following steps:

$$\widetilde{\mathbb{E}}Y = \mathbb{E}\big[ZY\big] = \mathbb{E}\big[\mathbb{E}_n[ZY]\big] = \mathbb{E}\big[Y\mathbb{E}_n[Z]\big] = \mathbb{E}\big[YZ_n\big]. \qquad \square$$

An illuminating application of Lemma 3.2.5 occurs if we fix a sequence of n coin tosses, $\overline{\omega}_1 \dots \overline{\omega}_n$, and define

$$Y(\omega_1 \dots \omega_n \omega_{n+1} \dots \omega_N) = \begin{cases} 1, & \text{if } \omega_1 \dots \omega_n = \overline{\omega}_1 \dots \overline{\omega}_n, \\ 0, & \text{otherwise.} \end{cases}$$

In other words, Y takes the value 1 if and only if the first n coin tosses result in the particular sequence $\overline{\omega}_1 \dots \overline{\omega}_n$ we have fixed in advance. The coin tosses $\omega_{n+1} \dots \omega_N$ are irrelevant. Then

$$\begin{aligned} \widetilde{\mathbb{E}}Y &= \widetilde{\mathbb{P}}\{\text{The first } n \text{ coin tosses result in } \overline{\omega}_1 \dots \overline{\omega}_n\} \\ &= \tilde{p}^{\#H(\overline{\omega}_1 \dots \overline{\omega}_n)} \tilde{q}^{\#T(\overline{\omega}_1 \dots \overline{\omega}_n)}, \end{aligned}$$

where the notation $\#H(\cdots)$ and $\#T(\cdots)$ is explained in Definition 3.1.3. On the other hand,

$$\begin{aligned} \mathbb{E}[YZ_n] &= Z_n(\overline{\omega}_1 \dots \overline{\omega}_n)\mathbb{P}\{\text{The first } n \text{ coin tosses result in } \overline{\omega}_1 \dots \overline{\omega}_n\} \\ &= Z_n(\overline{\omega}_1 \dots \overline{\omega}_n) p^{\#H(\overline{\omega}_1 \dots \overline{\omega}_n)} q^{\#T(\overline{\omega}_1 \dots \overline{\omega}_n)}. \end{aligned}$$

Lemma 3.2.5 asserts that these two quantities are equal, and hence

$$Z_n(\overline{\omega}_1 \dots \overline{\omega}_n) = \left(\frac{\tilde{p}}{p}\right)^{\#H(\overline{\omega}_1 \dots \overline{\omega}_n)} \left(\frac{\tilde{q}}{q}\right)^{\#T(\overline{\omega}_1 \dots \overline{\omega}_n)}. \tag{3.2.4}$$

This can be verified in Figure 3.2.1. For example, in that figure we have

$$Z_2(HH) = \left(\frac{\tilde{p}}{p}\right)^2 \left(\frac{\tilde{q}}{q}\right)^0 = \left(\frac{\frac{1}{2}}{\frac{2}{3}}\right)^2 = \frac{9}{16}$$

and

$$Z_3(HTT) = \left(\frac{\tilde{p}}{p}\right)^1 \left(\frac{\tilde{q}}{q}\right)^2 = \left(\frac{\frac{1}{2}}{\frac{2}{3}}\right)^1 \left(\frac{\frac{1}{2}}{\frac{1}{3}}\right)^2 = \frac{27}{16}.$$

We see that, for each n, $Z_n(\omega_1 \dots \omega_n)$ is the ratio of the risk-neutral probability and the actual probability of obtaining the sequence of coin tosses $\omega_1 \dots \omega_n$. Lemma 3.2.5 asserts that if Y depends only on the first n coin tosses, then we do not need to consider the coin tosses after time n. We may use Z_n as a surrogate for the Radon-Nikodým derivative Z in the formula $\widetilde{\mathbb{E}}Y = \mathbb{E}[ZY]$ of Theorem 3.1.1(iii), and Z_n is computed just like Z, except that Z_n is a ratio of probabilities for the first n coin tosses rather than all N tosses.

In addition to relating expectations under the two measures $\mathbb{P}$ and $\widetilde{\mathbb{P}}$, we want to have a formula relating conditional expectations under these measures. This is provided by the following lemma.

Lemma 3.2.6. *Assume the conditions of Definition 3.2.4. Let $n \le m$ be positive integers between 0 and N, and let Y be a random variable depending only on the first m coin tosses. Then*

$$\widetilde{\mathbb{E}}_n[Y] = \frac{1}{Z_n}\mathbb{E}_n[Z_m Y]. \tag{3.2.5}$$

PROOF: Let $\omega_1 \dots \omega_n$ be given. We compute

$$\begin{aligned} &\widetilde{\mathbb{E}}_n[Y](\omega_1 \dots \omega_n) \\ &= \sum_{\omega_{n+1} \dots \omega_m} Y(\omega_1 \dots \omega_m)\tilde{p}^{\#H(\omega_{n+1} \dots \omega_m)}\tilde{q}^{\#T(\omega_{n+1} \dots \omega_m)} \\ &= \left(\frac{p}{\tilde{p}}\right)^{\#H(\omega_1 \dots \omega_n)} \left(\frac{q}{\tilde{q}}\right)^{\#T(\omega_1 \dots \omega_n)} \\ &\quad \cdot \sum_{\omega_{n+1} \dots \omega_m} \Bigg[Y(\omega_1 \dots \omega_m) \left(\frac{\tilde{p}}{p}\right)^{\#H(\omega_1 \dots \omega_m)} \left(\frac{\tilde{q}}{q}\right)^{\#T(\omega_1 \dots \omega_m)} \\ &\qquad \cdot p^{\#H(\omega_{n+1} \dots \omega_m)} q^{\#T(\omega_{n+1} \dots \omega_m)} \Bigg] \\ &= \frac{1}{Z(\omega_1 \dots \omega_n)} \\ &\quad \cdot \sum_{\omega_{n+1} \dots \omega_m} Y(\omega_1 \dots \omega_m) Z_m(\omega_1 \dots \omega_m) p^{\#H(\omega_{n+1} \dots \omega_m)} q^{\#T(\omega_{n+1} \dots \omega_m)} \\ &= \frac{1}{Z(\omega_1 \dots \omega_n)}\mathbb{E}_n[YZ_m](\omega_1 \dots \omega_n). \qquad \square \end{aligned}$$

We are now in a position to give a variety of formulations of the risk-neutral pricing formula.

Theorem 3.2.7. *Consider an N-period binomial model with $0 < d < 1 + r < u$. Assume that the actual probability for head, p, and the actual probability for tail, q, are positive. The risk-neutral probabilities for head and tail are given, as usual, by*

$$\tilde{p} = \frac{1+r-d}{u-d}, \quad \tilde{q} = \frac{u-1-r}{u-d},$$

and these also are both positive. Let $\mathbb{P}$ *and* $\widetilde{\mathbb{P}}$ *denote the corresponding actual and risk-neutral probability measures, respectively, let* Z *be the Radon-Nikodým derivative of* $\widetilde{\mathbb{P}}$ *with respect to* $\mathbb{P}$, *and let* Z_n, $n = 0, 1, \ldots, N$, *be the Radon-Nikodým derivative process.*

Consider a derivative security whose payoff V_N *may depend on all* N *coin tosses. For* $n = 0, 1, \ldots, N$, *the price at time* n *of this derivative security is*

$$V_n = \widetilde{\mathbb{E}}_n \frac{V_N}{(1+r)^{N-n}} = \frac{(1+r)^n}{Z_n}\mathbb{E}_n \frac{Z_N V_N}{(1+r)^N} = \frac{1}{\zeta_n}\mathbb{E}_n[\zeta_N V_N], \tag{3.2.6}$$

where the state price density process ζ_n *is defined by*

$$\zeta_n = \frac{Z_n}{(1+r)^n}, \quad n = 0, 1, \ldots, N. \tag{3.2.7}$$

PROOF: The first equality in (3.2.6) is (2.4.11) of Chapter 2. The second equality follows from Lemma 3.2.6. The third is just a matter of definition of ζ_n. □

3.3 Capital Asset Pricing Model

The no-arbitrage pricing methodology of this text is one of two different ways of modeling prices of assets. The other, the *capital asset pricing model*, is based on balancing supply with demand among investors who have utility functions that convert units of consumption to units of satisfaction. The capital asset pricing model provides useful qualitative insights into markets but does not yield the precise quantitative results available through the no-arbitrage methodology. Moreover, in an idealized complete market, the no-arbitrage argument is compelling. On the other hand, many markets are incomplete, and prices cannot be determined from no-arbitrage considerations alone. Utility-based models are still the only theoretically defensible way of treating such markets, although there is a widespread practice of using "risk-neutral" pricing, even when the assets being priced cannot be replicated by trading in other, more primitive assets.

This text is about no-arbitrage pricing in complete markets and the mathematical methodology that supports this point of view. The mathematical methodology, however, is broadly applicable. In this section, we show how it can be brought to bear on a problem at the heart of the capital asset pricing model, that of maximizing the expected utility obtained from investment.

In no-arbitrage pricing, there are two probability measures, the actual probability measure and the risk-neutral measure. When pricing derivative securities, we need only consider the risk-neutral measure. There are, however, two situations in which the actual probability measure becomes relevant: asset management and risk management. In asset management, one cares about the trade-off of risk and actual (rather than risk-neutral) expected return. In risk management, one cares about the actual probability of a catastrophic event. In both of these situations, however, there is a role for the risk-neutral probability measure. For risk management, the portfolio whose risk is being assessed normally contains derivative securities whose theoretical prices under various scenarios must be computed using the risk-neutral measure. For asset management, the risk-neutral measure enters in the manner set forth in this section.

We now set out the capital asset pricing problem. By a *utility function* we shall mean a nondecreasing, concave function defined on the set of real numbers. This function may take the value $-\infty$, but not the value $+\infty$. A common utility function is $\ln x$, which is normally defined only for $x > 0$. We adopt the convention that $\ln x = -\infty$ for $x \le 0$, so this is defined for every $x \in \mathbb{R}$ and is nondecreasing and concave. Recall that a function U is concave if

$$U(\alpha x + (1-\alpha)y) \ge \alpha U(x) + (1-\alpha)U(y) \quad \text{for every } x, y \in \mathbb{R},\ \alpha \in (0,1). \tag{3.3.1}$$

We say U is *strictly concave* if the inequality in (3.3.1) is strict whenever $x \ne y$, and in fact we shall assume that U is strictly concave everywhere it is finite. A whole class of utility functions can be obtained by first choosing a number $p < 1$, $p \ne 0$, and another number $c \in \mathbb{R}$, and defining

$$U_p(x) = \begin{cases} \frac{1}{p}(x-c)^p, & \text{if } x > c, \\ 0, & \text{if } 0 < p < 1 \text{ and } x = c, \\ -\infty, & \text{if } p < 0 \text{ and } x = c, \\ -\infty, & \text{if } x < c. \end{cases}$$

For these functions, the *index of absolute risk aversion* $-\frac{U''(x)}{U'(x)}$ is the hyperbolic function $\frac{1-p}{x-c}$ for $x > c$. This class of functions is called the *HARA* (*h*yperbolic *a*bsolute *r*isk *a*version) class. The HARA function corresponding to $p = 0$ is

$$U_0(x) = \begin{cases} \ln(x-c), & \text{if } x > c, \\ -\infty, & \text{if } x \le c. \end{cases}$$

Concavity of utility functions is assumed in order to capture the trade-off between risk and return. For example, consider a gamble which pays off 1 with probability $\frac{1}{2}$ and 99 with probability $\frac{1}{2}$. The expected payoff is 50, but a risk-averse agent would prefer to have 50 rather than the random payoff of the gamble. Let X denote this random payoff, i.e., $\mathbb{P}(X=1) = \mathbb{P}(X=99) = \frac{1}{2}$. For a concave utility function, we have from Jensen's inequality used upside down (Theorem 2.2.5 of Chapter 2) that $\mathbb{E}U(X) \le U(\mathbb{E}X)$. Indeed, if $U(x) = \ln x$, then $\mathbb{E}\ln X = 2.30$ and $\ln \mathbb{E}X = 3.91$. If we model agent behavior as maximization of expected utility of payoff, our model would indicate that an agent would prefer the nonrandom payoff $\mathbb{E}X = 50$ over the random payoff X.

By comparing expected utility of payoffs rather than expected payoffs, and choosing the utility function judiciously, we can capture an investor's attitude toward the trade-off between risk and return.

Let us consider an N-period binomial model with the usual parameters $0 < d < 1+r < u$. An agent begins with initial wealth X_0 and wishes to invest in the stock and the money market account so as to maximize the expected utility of his wealth at time N. In other words, the agent has a utility function U and wishes to solve the following problem.

Problem 3.3.1 (Optimal investment). Given X_0, find an adapted portfolio process $\Delta_0, \Delta_1, \dots, \Delta_{N-1}$ that maximizes

$$\mathbb{E}U(X_N) \tag{3.3.2}$$

subject to the wealth equation

$$X_{n+1} = \Delta_n S_{n+1} + (1+r)(X_n - \Delta_n S_n), \; n = 0, 1, \dots, N-1.$$

Note that the expectation in (3.3.2) is computed using the actual probability measure $\mathbb{P}$. The agent is risk-averse and uses his utility function U to capture the trade-off between actual risk and actual return. It does not make sense to do this under the risk-neutral measure because under the risk-neutral measure both the stock and the money market account have the same rate of return; an agent seeking to maximize $\widetilde{\mathbb{E}}U(X)$ would invest only in the money market.

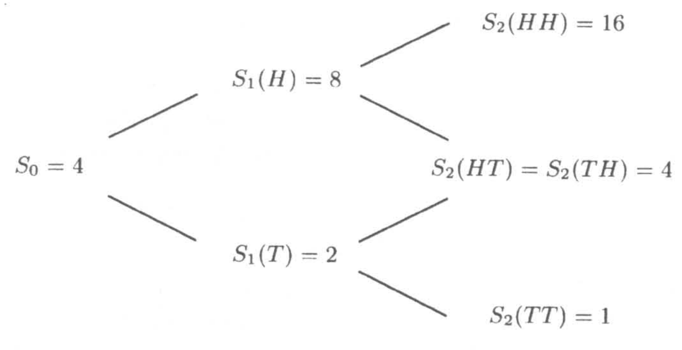

Fig. 3.3.1. A two-period model.

Example 3.3.2. Consider the two-period model of Figure 3.3.1, in which the interest rate is $r = \frac{1}{4}$, so the risk-neutral probability measure is

$$\widetilde{\mathbb{P}}(HH) = \frac{1}{4}, \; \widetilde{\mathbb{P}}(HT) = \frac{1}{4}, \; \widetilde{\mathbb{P}}(TH) = \frac{1}{4}, \; \widetilde{\mathbb{P}}(TT) = \frac{1}{4}.$$

Assume the actual probability measure is

$$\mathbb{P}(HH) = \frac{4}{9}, \; \mathbb{P}(HT) = \frac{2}{9}, \; \mathbb{P}(TH) = \frac{2}{9}, \; \mathbb{P}(TT) = \frac{1}{9}.$$

Consider an agent who begins with $X_0 = 4$ and wants to choose Δ_0, $\Delta_1(H)$ and $\Delta_1(T)$ in order to maximize $\mathbb{E}\ln X_2$. Note that

$$X_1(H) = 8\Delta_0 + \frac{5}{4}(4 - 4\Delta_0) = 3\Delta_0 + 5,$$

$$X_1(T) = 2\Delta_0 + \frac{5}{4}(4 - 4\Delta_0) = -3\Delta_0 + 5,$$

and

$$\begin{aligned}
X_2(HH) &= 16\Delta_1(H) + \frac{5}{4}(X_1(H) - 8\Delta_1(H)) \\
&= 6\Delta_1(H) + \frac{15}{4}\Delta_0 + \frac{25}{4}, && (3.3.3)\\
X_2(HT) &= 4\Delta_1(H) + \frac{5}{4}(X_1(H) - 8\Delta_1(H)) \\
&= -6\Delta_1(H) + \frac{15}{4}\Delta_0 + \frac{25}{4}, && (3.3.4)\\
X_2(TH) &= 4\Delta_1(T) + \frac{5}{4}(X_1(T) - 2\Delta_1(T)) \\
&= \frac{3}{2}\Delta_1(T) - \frac{15}{4}\Delta_0 + \frac{25}{4}, && (3.3.5)\\
X_2(TT) &= \Delta_1(T) + \frac{5}{4}(X_1(T) - 2\Delta_1(T)) \\
&= -\frac{3}{2}\Delta_1(T) - \frac{15}{4}\Delta_0 + \frac{25}{4}. && (3.3.6)
\end{aligned}$$

Therefore,

$$\begin{aligned}
&\mathbb{E}\ln X_2 \\
&= \frac{4}{9}\ln\left(6\Delta_1(H) + \frac{15}{4}\Delta_0 + \frac{25}{4}\right) + \frac{2}{9}\ln\left(-6\Delta_1(H) + \frac{15}{4}\Delta_0 + \frac{25}{4}\right) \\
&\quad + \frac{2}{9}\ln\left(\frac{3}{2}\Delta_1(T) - \frac{15}{4}\Delta_0 + \frac{25}{4}\right) + \frac{1}{9}\ln\left(-\frac{3}{2}\Delta_1(T) - \frac{15}{4}\Delta_0 + \frac{25}{4}\right).
\end{aligned}$$

The goal is to maximize this last expression. Toward this end, we compute the partial derivatives

$$\begin{aligned}\frac{\partial}{\partial \Delta_0}\mathbb{E}\ln X_2 &= \frac{4}{9}\cdot\frac{15}{4}\cdot\frac{1}{X_2(HH)}+\frac{2}{9}\cdot\frac{15}{4}\cdot\frac{1}{X_2(HT)}\\ &\quad -\frac{2}{9}\cdot\frac{15}{4}\cdot\frac{1}{X_2(TH)}-\frac{1}{9}\cdot\frac{15}{4}\cdot\frac{1}{X_2(TT)}\\ &= \frac{5}{12}\left(\frac{4}{X_2(HH)}+\frac{2}{X_2(HT)}-\frac{2}{X_2(TH)}-\frac{1}{X_2(TT)}\right),\\ \frac{\partial}{\partial \Delta_1(H)}\mathbb{E}\ln X_2 &= \frac{4}{9}\cdot\frac{6}{X_2(HH)}-\frac{2}{9}\cdot\frac{6}{X_2(HT)}\\ &= \frac{4}{3}\left(\frac{2}{X_2(HH)}-\frac{1}{X_2(HT)}\right),\\ \frac{\partial}{\partial \Delta_1(T)}\mathbb{E}\ln X_2 &= \frac{2}{9}\cdot\frac{3}{2}\cdot\frac{1}{X_2(TH)}-\frac{1}{9}\cdot\frac{3}{2}\cdot\frac{1}{X_2(TT)}\\ &= \frac{1}{6}\left(\frac{2}{X_2(TH)}-\frac{1}{X_2(TT)}\right).\end{aligned}$$

Setting these derivatives equal to zero, we obtain the three equations

$$\frac{4}{X_2(HH)}+\frac{2}{X_2(HT)}=\frac{2}{X_2(TH)}+\frac{1}{X_2(TT)}, \tag{3.3.7}$$

$$\frac{2}{X_2(HH)}=\frac{1}{X_2(HT)}, \tag{3.3.8}$$

$$\frac{2}{X_2(TH)}=\frac{1}{X_2(TT)}. \tag{3.3.9}$$

We can cross multiply in (3.3.8) and (3.3.9) to obtain

$$X_2(HH)=2X_2(HT), \tag{3.3.10}$$

$$X_2(TH)=2X_2(TT). \tag{3.3.11}$$

Substituting these equations into (3.3.7) and again cross multiplying, we obtain a third equation:

$$X_2(HT)=2X_2(TT). \tag{3.3.12}$$

This gives us the three linear equations (3.3.10)–(3.3.12) in the four unknowns $X_2(HH)$, $X_2(HT)$, $X_2(TH)$, and $X_2(TT)$.

One way to conclude is to recall the formulas (3.3.3)–(3.3.6) for $X_2(HH)$, $X_2(HT)$, $X_2(TH)$, and $X_2(TT)$ in terms of the three unknowns $\Delta_1(H)$, $\Delta_1(T)$, and Δ_0, substitute, and solve the resulting three linear equations in three unknowns. This will lead to the solutions

$$\Delta_0=\frac{5}{9},\quad \Delta_1(H)=\frac{25}{54},\quad \Delta_1(T)=\frac{25}{27}. \tag{3.3.13}$$

We have found the optimal portfolio, but the method we have used is not very pleasant. In particular, as the number of periods increases, the number

of variables $\Delta_n(\omega)$ grows exponentially, and in the last step we solved a system of linear equations in these variables.

An alternative way to conclude is to seek a fourth equation involving $X_2(HH)$, $X_2(HT)$, $X_2(TH)$, and $X_2(TT)$ to go with the three equations (3.3.10)–(3.3.12) and then solve these four equations in four unknowns. Such a fourth equation is provided by Corollary 2.4.6 of Chapter 2, which in this context says

$$4=\frac{16}{25}\left[\frac{1}{4}X_2(HH)+\frac{1}{4}X_2(HT)+\frac{1}{4}X_2(TH)+\frac{1}{4}X_2(TT)\right]. \tag{3.3.14}$$

It is now a straightforward matter to solve (3.3.10)–(3.3.12) and (3.3.14) to obtain

$$X_2(HH)=\frac{100}{9},\ X_2(HT)=\frac{50}{9},\ X_2(TH)=\frac{50}{9},\ X_2(TT)=\frac{25}{9}. \tag{3.3.15}$$

We can then find $\Delta_1(H)$, $\Delta_1(T)$, and Δ_0 by the algorithm of Theorem 1.2.2 of Chapter 1. In particular,

$$\begin{aligned}\Delta_1(H)&=\frac{X_2(HH)-X_2(HT)}{S_2(HH)-S_2(HT)} &&=\frac{25}{54},\\ \Delta_1(T)&=\frac{X_2(TH)-X_2(TT)}{S_2(TH)-S_2(TT)} &&=\frac{25}{27},\\ X_1(H)&=\frac{4}{5}\left[\frac{1}{2}X_2(HH)+\frac{1}{2}X_2(HT)\right] &&=\frac{20}{3},\\ X_1(T)&=\frac{4}{5}\left[\frac{1}{2}X_2(TH)+\frac{1}{2}X_2(TT)\right] &&=\frac{10}{3},\\ \Delta_0&=\frac{X_1(H)-X_1(T)}{S_1(H)-S_1(T)} &&=\frac{5}{9}.\end{aligned} \tag{3.3.16}$$

□

The second method of concluding the preceding example used equation (3.3.14), which follows from the fact that the expected discounted value of a portfolio process under the risk-neutral measure is always equal to the initial value X_0 (see Corollary 2.4.6 of Chapter 2). In general,

$$\widetilde{\mathbb{E}}\frac{X_N}{(1+r)^N}=X_0. \tag{3.3.17}$$

This equation introduces the risk-neutral measure to the solution of Problem 3.3.1, even though only the actual probability measure appears in the statement of the problem. This suggests that we might replace Problem 3.3.1 by the following problem.

Problem 3.3.3. Given X_0, find a random variable X_N (without regard to a portfolio process) that maximizes

$$\mathbb{E}U(X_N) \tag{3.3.18}$$

subject to

$$\widetilde{\mathbb{E}}\frac{X_N}{(1+r)^N} = X_0. \tag{3.3.19}$$

Lemma 3.3.4. *Suppose $\Delta_0^*, \Delta_1^*, \ldots, \Delta_{N-1}^*$ is an optimal portfolio process for Problem 3.3.1, and X_N^* is the corresponding optimal wealth random variable at time N. Then X_N^* is optimal for Problem 3.3.3. Conversely, suppose X_N^* is optimal for Problem 3.3.3. Then there is a portfolio process $\Delta_0^*, \Delta_1^*, \ldots, \Delta_{N-1}^*$ that starts with initial wealth X_0 and has value X_N^* at time N, and this portfolio process is optimal for Problem 3.3.1.*

PROOF: Assume first that $\Delta_0^*, \Delta_1^*, \ldots, \Delta_{N-1}^*$ is an optimal portfolio process for Problem 3.3.1, and X_N^* is the corresponding optimal wealth random variable at time N. To show that X_N^* is optimal for Problem 3.3.3, we must show that it satisfies the constraint (3.3.19) and that $\mathbb{E}U(X_N) \le \mathbb{E}U(X_N^*)$ for any other X_N that satisfies this constraint. Because it is generated by a portfolio starting with initial wealth X_0, the random variable X_N^* satisfies (3.3.17), which is (3.3.19). Now let X_N be any other random variable satisfying (3.3.19). We may regard X_N as a derivative security, and according to the risk-neutral pricing formula (2.4.11) of Chapter 2, the time-zero price of this derivative security is X_0 appearing in (3.3.19). In particular, beginning with initial wealth X_0, we may construct a portfolio process $\Delta_1, \Delta_2, \ldots, \Delta_{N-1}$ that replicates X_N (i.e., for which the value of the portfolio process at time N is X_N). (See Theorem 1.2.2 of Chapter 1 for the details.) Since X_N^* is an optimal final portfolio random variable for Problem 3.3.1 and X_N is another final portfolio random variable, we must have $\mathbb{E}U(X_N) \le \mathbb{E}U(X_N^*)$. This shows that X_N^* is optimal for Problem 3.3.3.

For the converse, suppose X_N^* is optimal for Problem 3.3.3. Again using Theorem 1.2.2 of Chapter 1, we may construct a portfolio process $\Delta_0^*, \Delta_1^*, \ldots, \Delta_{N-1}^*$ that begins with initial wealth X_0 and whose value at time N is X_N^*. Let $\Delta_0, \Delta_1, \ldots, \Delta_{N-1}$ be another portfolio process, which, starting with initial wealth X_0, leads to some wealth X_N at time N. To show that X_N^* is optimal in Problem 3.3.1, we must show that

$$\mathbb{E}U(X_N) \le \mathbb{E}U(X_N^*). \tag{3.3.20}$$

But X_N satisfies (3.3.17), which is (3.3.19), and X_N^* is optimal for Problem 3.3.3. This implies (3.3.20) and establishes the optimality of X_N^* in Problem 3.3.1. □

Lemma 3.3.4 separates the optimal investment problem, Problem 3.3.1, into two manageable steps: first, find a random variable X_N that solves Problem 3.3.3; and second, construct the portfolio that starts with X_0 and replicates X_N. The second step uses the algorithm of Theorem 1.2.2 of Chapter 1. It remains only to figure out how to perform the first step. Before giving the general method, we examine Problem 3.3.1 within the context of Example 3.3.2.

Example 3.3.2 (continued) We first compute the Radon-Nikodým derivative of $\widetilde{\mathbb{P}}$ with respect to $\mathbb{P}$:

$$Z(HH) = \frac{\widetilde{\mathbb{P}}(HH)}{\mathbb{P}(HH)} = \frac{9}{16}, \quad Z(HT) = \frac{\widetilde{\mathbb{P}}(HT)}{\mathbb{P}(HT)} = \frac{9}{8},$$
$$Z(TH) = \frac{\widetilde{\mathbb{P}}(TH)}{\mathbb{P}(TH)} = \frac{9}{8}, \quad Z(TT) = \frac{\widetilde{\mathbb{P}}(TT)}{\mathbb{P}(TT)} = \frac{9}{4}.$$

To simplify matters, we use subscripts to denote the different values of the state price density ζ:

$$\zeta_1 = \zeta(HH) = \frac{Z(HH)}{(1+r)^2} = \frac{9}{25},$$
$$\zeta_2 = \zeta(HT) = \frac{Z(HT)}{(1+r)^2} = \frac{18}{25},$$
$$\zeta_3 = \zeta(TH) = \frac{Z(TH)}{(1+r)^2} = \frac{18}{25},$$
$$\zeta_4 = \zeta(TT) = \frac{Z(TT)}{(1+r)^2} = \frac{36}{25}.$$

We also use the notation

$$p_1 = \mathbb{P}(HH) = \frac{4}{9}, \quad p_2 = \mathbb{P}(HT) = \frac{2}{9},$$
$$p_3 = \mathbb{P}(TH) = \frac{2}{9}, \quad p_4 = \mathbb{P}(TT) = \frac{1}{9}.$$

Finally, we denote

$$x_1 = X_2(HH), \quad x_2 = X_2(HT),$$
$$x_3 = X_2(TH), \quad x_4 = X_2(TT).$$

With these notations, Problem 3.3.3 may be written as

Find a vector (x_1, x_2, x_3, x_4) that maximizes $\sum_{m=1}^4 p_m U(x_m)$ subject to $\sum_{m=1}^4 p_m \zeta_m x_m = X_0$.

Filling in the numbers and using the fact that the utility function in question is the logarithm, we rewrite this as

Find a vector (x_1, x_2, x_3, x_4) that maximizes

$$\frac{4}{9}\ln x_1 + \frac{2}{9}\ln x_2 + \frac{2}{9}\ln x_3 + \frac{1}{9}\ln x_4$$

subject to

$$\frac{4}{9}\cdot\frac{9}{25}x_1+\frac{2}{9}\cdot\frac{18}{25}x_2+\frac{2}{9}\cdot\frac{18}{25}x_3+\frac{1}{9}\cdot\frac{36}{25}x_4=4. \tag{3.3.21}$$

The Lagrangian for this problem is

$$\begin{aligned}L&=\frac{4}{9}\ln x_1+\frac{2}{9}\ln x_2+\frac{2}{9}\ln x_3+\frac{1}{9}\ln x_4\\&\quad-\lambda\left(\frac{4}{9}\cdot\frac{9}{25}x_1+\frac{2}{9}\cdot\frac{18}{25}x_2+\frac{2}{9}\cdot\frac{18}{25}x_3+\frac{1}{9}\frac{36}{25}\cdot x_4-4\right).\end{aligned}$$

The Lagrange multiplier equations are

$$\begin{aligned}\frac{\partial}{\partial x_1}L&=\frac{4}{9}\left(\frac{1}{x_1}-\lambda\frac{9}{25}\right)=0,\\\frac{\partial}{\partial x_2}L&=\frac{2}{9}\left(\frac{1}{x_2}-\lambda\frac{18}{25}\right)=0,\\\frac{\partial}{\partial x_3}L&=\frac{2}{9}\left(\frac{1}{x_3}-\lambda\frac{18}{25}\right)=0,\\\frac{\partial}{\partial x_4}L&=\frac{1}{9}\left(\frac{1}{x_4}-\lambda\frac{36}{25}\right)=0,\end{aligned}$$

which imply

$$x_1=\frac{25}{9\lambda},\ x_2=\frac{25}{18\lambda},\ x_3=\frac{25}{18\lambda},\ x_4=\frac{25}{36\lambda}.$$

We solve for $\frac{1}{\lambda}$ by substituting these formulas into (3.3.21):

$$\frac{4}{9\lambda}+\frac{2}{9\lambda}+\frac{2}{9\lambda}+\frac{1}{9\lambda}=4,$$

which shows that $\frac{1}{\lambda}=4$. We conclude that the optimal wealth at time two is

$$\begin{aligned}X_2(HH)&=x_1=\frac{100}{9},\quad X_2(HT)=x_2=\frac{50}{9},\\X_2(TH)&=x_3=\frac{50}{9},\quad X_2(TT)=x_4=\frac{25}{9}.\end{aligned}$$

This agrees with formula (3.3.15). We can now compute the optimal portfolio process Δ_0, $\Delta_1(H)$, and $\Delta_1(T)$ as we did following that formula. □

In general, the solution of Problem 3.3.3 follows along the lines of the previous example. It is complicated by the fact that both the actual and risk-neutral probability measures appear in the problem formulation. Consequently, we introduce the Radon-Nikodým derivative Z of $\widetilde{\mathbb{P}}$ with respect to $\mathbb{P}$ to rewrite (3.3.19) without reference to the risk-neutral measure. This constraint becomes

$$\mathbb{E}\frac{Z_N X_N}{(1+r)^N}=X_0. \tag{3.3.19)$'$}$$

We can take this one step further by recalling the state price density $\zeta=\frac{Z}{(1+r)^N}$, in terms of which (3.3.19) can be written as

$$\mathbb{E}\zeta X_N=X_0. \tag{3.3.19)$''$}$$

In an N-period model, there are $M=2^N$ possible coin toss sequences ω. Let us list them, labeling them

$$\omega^1,\omega^2,\dots,\omega^M.$$

We use superscripts to indicate that ω^m is a full sequence of coin tosses, not the mth coin toss of some sequence. Let us define $\zeta_m=\zeta(\omega^m)$, $p_m=\mathbb{P}(\omega^m)$, and $x_m=X_N(\omega^m)$. Then Problem 3.3.3 can be reformulated as follows.

Problem 3.3.5. Given X_0, find a vector $(x_1,x_2,\dots,x_M)$ that maximizes

$$\sum_{m=1}^M p_m U(x_m)$$

subject to

$$\sum_{m=1}^M p_m x_m\zeta_m=X_0.$$

The Lagrangian for Problem 3.3.5 is

$$L=\sum_{m=1}^M p_m U(x_m)-\lambda\left(\sum_{m=1}^M p_m x_m\zeta_m-X_0\right),$$

and the Lagrange multiplier equations are

$$\frac{\partial}{\partial x_m}L=p_m U'(x_m)-\lambda p_m\zeta_m=0,\ m=1,2,\dots,M. \tag{3.3.22}$$

These equations reduce to

$$U'(x_m)=\lambda\zeta_m,\ m=1,2,\dots,M. \tag{3.3.23}$$

Recalling how x_m and ζ_m were defined, we rewrite this as

$$U'(X_N)=\frac{\lambda Z}{(1+r)^N}. \tag{3.3.24}$$

At this point, we need to invert the function U'. Since U is strictly concave everywhere it is finite, its derivative is decreasing and so has an inverse function, which we call I. For example, if $U(x)=\ln x$, then $U'(x)=\frac{1}{x}$. Setting

$y = U'(x) = \frac{1}{x}$, we solve for $x = \frac{1}{y}$, and this determines the inverse function $I(y) = \frac{1}{y}$. After determining this inverse function, whatever it is, we invert (3.3.24) to obtain

$$X_N = I\left(\frac{\lambda Z}{(1+r)^N}\right). \tag{3.3.25}$$

This gives a formula for the optimal X_N in terms of the multiplier λ. We solve for the multiplier λ by substituting X_N into (3.3.19)$'$:

$$\mathbb{E}\left[\frac{Z}{(1+r)^N} I\left(\frac{\lambda Z}{(1+r)^N}\right)\right] = X_0. \tag{3.3.26}$$

After solving this equation for λ, we substitute λ into (3.3.25) to obtain X_N, and then we use the algorithm in Theorem 1.2.2 of Chapter 1 to find the optimal portfolio process $\Delta_0, \Delta_1, \dots, \Delta_{N-1}$. All these steps were carried out in Example 3.3.2 (continued).

We summarize this discussion with a theorem.

Theorem 3.3.6. *The solution of Problem 3.3.1 can be found by first solving equation (3.3.26) for λ, then computing X_N by (3.3.25), and finally using X_N in the algorithm of Theorem 1.2.2 of Chapter 1 to determine the optimal portfolio process $\Delta_0, \Delta_1, \dots, \Delta_{N-1}$ and corresponding portfolio value process $X_1, X_2, \dots, X_N$. The function I appearing in (3.3.26) is the functional inverse of the derivative U' of the utility function U in Problem 3.3.1; i.e., $x = I(y)$ if and only if $y = U'(x)$.*

3.4 Summary

This chapter details the methodology for changing from the actual probability measure to the risk-neutral probability measure in a binomial model and, more generally, the methodology for changing from one probability measure to another in a finite probability model. The key quantity is the *Radon-Nikodým derivative*

$$Z(\omega) = \frac{\widetilde{\mathbb{P}}(\omega)}{\mathbb{P}(\omega)}, \tag{3.1.1}$$

which in a finite probability model is just the quotient of the two probability measures. The Radon-Nikodým derivative is a strictly positive random variable with $\mathbb{E}Z = 1$. The expectations of a random variable Y under the two probability measures are related by the formula

$$\widetilde{\mathbb{E}}Y = \mathbb{E}[ZY]. \tag{3.1.2}$$

In the binomial model with actual probabilities p and q and risk-neutral probabilities $\tilde{p}$ and $\tilde{q}$ for head and tail, respectively, in addition to the *Radon-Nikodým derivative random variable*

$$Z(\omega_1 \dots \omega_N) = \frac{\widetilde{\mathbb{P}}(\omega_1 \dots \omega_N)}{\mathbb{P}(\omega_1 \dots \omega_N)} = \left(\frac{\tilde{p}}{p}\right)^{\#H(\omega_1 \dots \omega_N)} \left(\frac{\tilde{q}}{q}\right)^{\#T(\omega_1 \dots \omega_N)}, \tag{3.1.8}$$

we have a *Radon-Nikodým derivative process*

$$Z_n = \mathbb{E}_n Z, \; n = 0, 1, \dots, N. \tag{3.2.1}$$

This process is also given by the formula

$$Z_n(\omega_1 \dots \omega_n) = \left(\frac{\tilde{p}}{p}\right)^{\#H(\omega_1 \dots \omega_n)} \left(\frac{\tilde{q}}{q}\right)^{\#T(\omega_1 \dots \omega_n)}. \tag{3.2.4}$$

In other words, $Z_n(\omega_1 \dots \omega_n)$ is the ratio of the risk-neutral probability of the partial path of n tosses to the actual probability of the same partial path. When the random variable Y depends only on the first n tosses, where $0 \le n \le N$, equation (3.1.2) takes the simpler form

$$\widetilde{\mathbb{E}}Y = \mathbb{E}[Z_n Y]. \tag{3.2.3}$$

This shows that when Y is determined by the outcome of the first n coin tosses, then we need only consider the ratio of the risk-neutral probability to the actual probability for these n tosses in order to relate $\widetilde{\mathbb{P}}$ and $\mathbb{P}$ expectations of Y.

Conditional expectations under $\widetilde{\mathbb{P}}$ and $\mathbb{P}$ are related as follows. If Y depends only on the first m coin tosses and $0 \le n \le m \le N$, then

$$\widetilde{\mathbb{E}}_n Y = \frac{1}{Z_n} \mathbb{E}_n[Z_m Y]. \tag{3.2.5}$$

When computing the conditional expectation $\widetilde{\mathbb{E}}_n Y$, we imagine we have seen the coin tosses $\omega_1 \dots \omega_n$ and we have assumed that Y does not depend on the tosses $\omega_{m+1} \dots \omega_N$. The coin tosses $\omega_{n+1} \dots \omega_m$, which we have not seen and affect the value of Y, have $\widetilde{\mathbb{P}}$-probability $\tilde{p}^{\#H(\omega_{n+1} \dots \omega_m)} \tilde{q}^{\#T(\omega_{n+1} \dots \omega_m)}$ and $\mathbb{P}$-probability $p^{\#H(\omega_{n+1} \dots \omega_m)} q^{\#T(\omega_{n+1} \dots \omega_m)}$. The ratio of these two probabilities is

$$\frac{Z_m(\omega_1 \dots \omega_m)}{Z_n(\omega_1 \dots \omega_n)} = \left(\frac{\tilde{p}}{p}\right)^{\#H(\omega_{n+1} \dots \omega_m)} \left(\frac{\tilde{q}}{q}\right)^{\#T(\omega_{n+1} \dots \omega_m)},$$

and thus this quotient random variable is used to write the $\widetilde{\mathbb{P}}$-conditional expectation in terms of the $\mathbb{P}$-conditional expectation in (3.2.5). Note in this regard that the right-hand side of (3.2.5) may also be written as $\mathbb{E}_n\left[\frac{Z_m}{Z_n} Y\right]$ since Z_n depends only on the first n coin tosses.

The Radon-Nikodým derivative random variable Z in the binomial model gives rise to the *state price density*

$$\zeta(\omega) = \frac{Z(\omega)}{(1+r)^N}.$$

We may interpret $\zeta(\omega)$ as the value at time zero per unit of actual probability of a derivative security that pays 1 at time N if the coin tossing results in the sequence ω (see the discussion following Definition 3.1.3). In other words, $\zeta(\omega)\mathbb{P}(\omega)$, the so-called *state price* of ω, is the value at time zero of a derivative security that pays 1 at time N if the coin tossing results in the sequence ω.

We can use the state price density to solve the following optimal investment problem.

Problem 3.3.1 (Optimal investment) Given X_0, find an adapted portfolio process $\Delta_0, \Delta_1, \ldots, \Delta_{N-1}$ that maximizes

$$\mathbb{E}U(X_N), \tag{3.3.2}$$

subject to the wealth equation

$$X_{n+1} = \Delta_n S_{n+1} + (1+r)(X_n - \Delta_n S_n), \quad n = 0, 1, \ldots, N-1.$$

If we list the $M = 2^N$ possible coin tosses sequences ω in the N-period binomial model, labeling them $\omega^1, \omega^2, \ldots, \omega^M$, and if we define $\zeta_m = \zeta(\omega^m)$, $p_m = \mathbb{P}(\omega^m)$, and $x_m = X_N(\omega^m)$, then Problem 3.3.1 may be reduced to the following problem.

Problem 3.3.5 Given X_0, find a vector $(x_1, x_2, \ldots, x_M)$ that maximizes

$$\sum_{m=1}^{M} p_m U(x_m)$$

subject to

$$\sum_{m=1}^{M} p_m x_m \zeta_m = X_0.$$

In Problem 3.3.1, the search is over all portfolio processes, whereas in Problem 3.3.5 the search is over the M variables $x_1, \ldots, x_M$. The second problem is simpler and can be solved by the method of Lagrange multipliers. The optimal values of x_m satisfy

$$U'(x_m) = \lambda\zeta_m, \quad m = 1, 2, \ldots, M, \tag{3.3.23}$$

where λ is the Lagrange multiplier. This leads to the formula

$$X_N = I\left(\frac{\lambda Z}{(1+r)^N}\right), \tag{3.3.25}$$

where I is the inverse of the strictly decreasing function U' and the Lagrange multiplier λ is chosen so that the equation

$$\mathbb{E}\left[\frac{Z}{(1+r)^N} I\left(\frac{\lambda Z}{(1+r)^N}\right)\right] = X_0 \tag{3.3.26}$$

holds. Once the optimal terminal wealth X_N is determined by these equations, we treat it as if it were the payoff of a derivative security and determine the portfolio process $\Delta_0, \Delta_1, \ldots, \Delta_{N-1}$ which solves Problem 3.3.1 as the hedge for a short position in this derivative security, using the algorithm of Theorem 1.2.2 of Chapter 1.

3.5 Notes

One might think that the value at time zero of a payoff of 1 at time N if the coin tosses result in "state" ω should be just the expected payoff, which is the (actual rather than risk-neutral) probability of ω. However, this does not account for the risk that the payoff is zero if "state" omega does not occur. The creation of a *state price* to account for risk traces back to Arrow and Debreu [1].

The problem of optimal investment or, more generally, optimal consumption and investment, has been the subject of a great deal of research. The original papers are Hakansson [16] for the discrete-time model and Merton [32], [33], [35] for the continuous-time model. (These papers of Merton are collected in [36].) The solution of the problem via the state price density process is due to Pliska [37] and was further developed by Cox and Huang [7], [8] and Karatzas, Lehoczky, and Shreve [27]. Except for [16], all these models are in continuous time. They have been specialized to the binomial model for this text. A compilation of research along these lines, including treatment of portfolio constraints and the model with different interest rates for borrowing and investment, is provided by Karatzas and Shreve [28].

3.6 Exercises

Exercise 3.1. Under the conditions of Theorem 3.1.1, show the following analogues of properties (i)–(iii) of that theorem:

(i′) $\widetilde{\mathbb{P}}\left(\frac{1}{Z} > 0\right) = 1$;
(ii′) $\widetilde{\mathbb{E}}\frac{1}{Z} = 1$;
(iii′) for any random variable Y,

$$\mathbb{E}Y = \widetilde{\mathbb{E}}\left[\frac{1}{Z} \cdot Y\right].$$

In other words, $\frac{1}{Z}$ facilitates the switch from $\widetilde{\mathbb{E}}$ to $\mathbb{E}$ in the same way Z facilitates the switch from $\mathbb{E}$ to $\widetilde{\mathbb{E}}$.

Exercise 3.2. Let $\mathbb{P}$ be a probability measure on a finite probability space Ω. In this problem, we allow the possibility that $\mathbb{P}(\omega) = 0$ for some values of $\omega \in \Omega$. Let Z be a random variable on Ω with the property that $\mathbb{P}(Z \geq 0) = 1$ and $\mathbb{E}Z = 1$. For $\omega \in \Omega$, define $\widetilde{\mathbb{P}}(\omega) = Z(\omega)\mathbb{P}(\omega)$, and for events $A \subset \Omega$, define $\widetilde{\mathbb{P}}(A) = \sum_{\omega \in A} \widetilde{\mathbb{P}}(\omega)$. Show the following.

(i) $\widetilde{\mathbb{P}}$ is a probability measure; i.e., $\widetilde{\mathbb{P}}(\Omega) = 1$.
(ii) If Y is a random variable, then $\widetilde{\mathbb{E}}Y = \mathbb{E}[ZY]$.
(iii) If A is an event with $\mathbb{P}(A) = 0$, then $\widetilde{\mathbb{P}}(A) = 0$.

(iv) Assume that $\mathbb{P}(Z > 0) = 1$. Show that if A is an event with $\widetilde{\mathbb{P}}(A) = 0$, then $\mathbb{P}(A) = 0$.

When two probability measures agree which events have probability zero (i.e., $\mathbb{P}(A) = 0$ if and only if $\widetilde{\mathbb{P}}(A) = 0$), the measures are said to be *equivalent*. From (iii) and (iv) above, we see that $\mathbb{P}$ and $\widetilde{\mathbb{P}}$ are equivalent under the assumption that $\mathbb{P}(Z > 0) = 1$.

(v) Show that if $\mathbb{P}$ and $\widetilde{\mathbb{P}}$ are equivalent, then they agree which events have probability one (i.e., $\mathbb{P}(A) = 1$ if and only if $\widetilde{\mathbb{P}}(A) = 1$).
(vi) Construct an example in which we have only $\mathbb{P}(Z \geq 0) = 1$ and $\mathbb{P}$ and $\widetilde{\mathbb{P}}$ are not equivalent.

In finance models, the risk-neutral probability measure and actual probability measure must always be equivalent. They agree about what is possible and what is impossible.

Exercise 3.3. Using the stock price model of Figure 3.1.1 and the actual probabilities $p = \frac{2}{3}$, $q = \frac{1}{3}$, define the estimates of S_3 at various times by

$$M_n = \mathbb{E}_n[S_3], \quad n = 0, 1, 2, 3.$$

Fill in the values of M_n in a tree like that of Figure 3.1.1. Verify that M_n, $n = 0, 1, 2, 3$, is a martingale.

Exercise 3.4. This problem refers to the model of Example 3.1.2, whose Radon-Nikodým process Z_n appears in Figure 3.2.1.

(i) Compute the state price densities

$$\begin{aligned}&\zeta_3(HHH),\\&\zeta_3(HHT) = \zeta_3(HTH) = \zeta_3(THH),\\&\zeta_3(HTT) = \zeta_3(THT) = \zeta_3(TTH),\\&\zeta_3(TTT)\end{aligned}$$

explicitly.
(ii) Use the numbers computed in (i) in formula (3.1.10) to find the time-zero price of the Asian option of Exercise 1.8 of Chapter 1. You should get $v_0(4, 4)$ computed in part (ii) of that exercise.
(iii) Compute also the state price densities $\zeta_2(HT) = \zeta_2(TH)$.
(iv) Use the risk-neutral pricing formula (3.2.6) in the form

$$\begin{aligned}V_2(HT) &= \frac{1}{\zeta_2(HT)}\mathbb{E}_2[\zeta_3 V_3](HT),\\V_2(TH) &= \frac{1}{\zeta_2(TH)}\mathbb{E}_2[\zeta_3 V_3](TH)\end{aligned}$$

to compute $V_2(HT)$ and $V_2(TH)$. You should get $V_2(HT) = v_2(4, 16)$ and $V_2(TH) = v_2(4, 10)$, where $v_2(s, y)$ was computed in part (ii) of Exercise 1.8 of Chapter 1. Note that $V_2(HT) \neq V_2(TH)$.

Exercise 3.5 (Stochastic volatility, random interest rate). Consider the model of Exercise 2.9 of Chapter 2. Assume that the actual probability measure is

$$\mathbb{P}(HH) = \frac{4}{9}, \ \mathbb{P}(HT) = \frac{2}{9}, \ \mathbb{P}(TH) = \frac{2}{9}, \ \mathbb{P}(TT) = \frac{1}{9}.$$

The risk-neutral measure was computed in Exercise 2.9 of Chapter 2.

(i) Compute the Radon-Nikodým derivative $Z(HH)$, $Z(HT)$, $Z(TH)$, and $Z(TT)$ of $\widetilde{\mathbb{P}}$ with respect to $\mathbb{P}$.
(ii) The Radon-Nikodým derivative process Z_0, Z_1, Z_2 satisfies $Z_2 = Z$. Compute $Z_1(H)$, $Z_1(T)$, and Z_0. Note that $Z_0 = \mathbb{E}Z = 1$.
(iii) The version of the risk-neutral pricing formula (3.2.6) appropriate for this model, which does not use the risk-neutral measure, is

$$\begin{aligned}V_1(H) &= \frac{1+r_0}{Z_1(H)}\mathbb{E}_1\left[\frac{Z_2}{(1+r_0)(1+r_1)}V_2\right](H)\\&= \frac{1}{Z_1(H)(1+r_1(H))}\mathbb{E}_1[Z_2V_2](H),\\V_1(T) &= \frac{1+r_0}{Z_1(T)}\mathbb{E}_1\left[\frac{Z_2}{(1+r_0)(1+r_1)}V_2\right](T)\\&= \frac{1}{Z_1(T)(1+r_1(T))}\mathbb{E}_1[Z_2V_2](T),\\V_0 &= \mathbb{E}\left[\frac{Z_2}{(1+r_0)(1+r_1)}V_2\right].\end{aligned}$$

Use this formula to compute $V_1(H)$, $V_1(T)$, and V_0 when $V_2 = (S_2 - 7)^+$. Compare the result with your answers in Exercise 2.6(ii) of Chapter 2.

Exercise 3.6. Consider Problem 3.3.1 in an N-period binomial model with the utility function $U(x) = \ln x$. Show that the optimal wealth process corresponding to the optimal portfolio process is given by $X_n = \frac{X_0}{\zeta_n}$, $n = 0, 1, \ldots, N$, where ζ_n is the state price density process defined in (3.2.7).

Exercise 3.7. Consider Problem 3.3.1 in an N-period binomial model with the utility function $U(x) = \frac{1}{p}x^p$, where $p < 1$, $p \neq 0$. Show that the optimal wealth at time N is

$$X_N = \frac{X_0(1+r)^N Z^{\frac{1}{p-1}}}{\mathbb{E}\left[Z^{\frac{p}{p-1}}\right]},$$

where Z is the Radon-Nikodým derivative of $\widetilde{\mathbb{P}}$ with respect to $\mathbb{P}$.

Exercise 3.8. The Lagrange Multiplier Theorem used in the solution of Problem 3.3.5 has hypotheses that we did not verify in the solution of that problem. In particular, the theorem states that if the gradient of the constraint function,